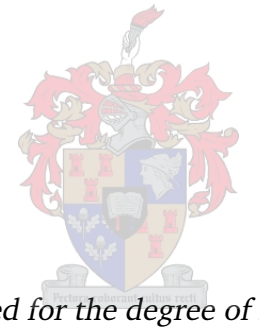
Development of an Integrated Metabolic Analysis Toolbox

by

Carl David Christensen



Dissertation presented for the degree of Doctor of Philosophy (Biochemistry) in the Faculty of Science at Stellenbosch University

Promoter: Prof. JM Rohwer

Co-promoter: Prof. J-HS Hofmeyr

December 2016

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: ..... December 2016 .....

Abstract

Development of an Integrated Metabolic Analysis Toolbox

C.D. Christensen

Department of Biochemistry,
Stellenbosch University,
Private Bag X1, Matieland 7602, South Africa.

Dissertation: PhD (Biochem)

December 2016

Life is arguably the most complex of all natural phenomena, yet it arises from essentially dead molecular components. The goal of systems biology is to be able to understand how the properties and non-linear interactions of these components give rise to the functions and behaviour of living biological systems. This represents the so-called "mechanistic explanation" where no individual component, nor the complete system itself, is privileged.

In this dissertation a Python based software package called `PySCeSToolbox` is presented that includes tools that implement previously published theoretical frameworks for investigating kinetic models of metabolic systems. These tools are `RateChar`, which performs generalised supply-demand analysis (GSDA); `SymCa`, which performs symbolic metabolic control analysis; and `ThermoKin`, which distinguishes between the kinetic and thermodynamic contributions towards enzyme-catalysed reaction rates. Each of the frameworks contained within the tools of `PySCeSToolbox` views metabolism from a different vantage point: generalised supply-demand analysis gives a broad overview of the behaviour, control, and regulation of metabolic systems by taking into account their functional organisation; symbolic control analysis dissects the control properties of metabolic systems in terms of the physical chains of interactions between enzymes and metabolic intermediates; and the thermodynamic/kinetic framework zooms in on the properties of the enzymes themselves to determine their regulatory roles. The strength of `PySCeSToolbox` lies in its integration of these viewpoints into a single analysis package in a way that promotes their complementary use in the search for a mechanistic explanation of modelled metabolic systems.

Through the application of these tools in the investigation of two previously published metabolic models, new knowledge regarding their behaviour is uncovered and subsequently explained in terms of their component properties and interactions. In a model of aspartate-derived amino-acid synthesis, a GSDA reveals that aspartate-semialdehyde regulates the reaction block that produces it via the reaction blocks that consume it, in spite of the relatively high sensitivity of its supply enzyme towards this intermediate. Subsequently, the regulatory contributions of each of the four aspartate-semialdehyde consuming blocks towards the producing block are quantified. In a model of pyruvate branch metabolism, application of GSDA shows that the flux through a NADH/NAD^+ consuming reaction block decreases when the ratio of NADH to NAD^+ increases. Rather than being a result of substrate inhibition, this phenomenon is shown to be the result of an interaction of the NADH/NAD^+ intermediates with a reaction elsewhere in the pathway.

Symbolic control analysis of the pyruvate branch model exposes a number of features that explain the unintuitive flux response described above. Firstly, only some control patterns are important for determining the flux control at any time. Secondly, different control patterns are dominant under different conditions, and dominance shifts as these conditions change. Finally, dissection of these chains of effects identifies the components of the system that are responsible for the flux control. Additional use of the thermodynamic/kinetic framework to focus on the enzymes that constitute the control patterns relates their values to the properties of individual enzyme-catalysed reactions (i.e. their elasticities). This framework is also used to explain the behaviour of the elasticity coefficient components of the unintuitive flux response, which are shown to be mostly mass-action controlled. Ultimately this two-pronged strategy provides a mechanistic explanation of the flux response, in which this high-level property is quantitatively linked to various low-level components.

The design of `PySCeSToolbox` as a Python-based software library allows it to integrate with the existing scientific Python ecosystem, thus providing access to a variety of additional third-party software tools to aid in the analysis of metabolic systems. This design also encourages the use of a scripting approach to designing *in silico* modelling experiments, which in turn promotes reproducibility through the re-use of such scripts. Moreover, `PySCeSToolbox` provides computational access to theoretical analysis frameworks that would otherwise have been inaccessible to researchers, as these frameworks are not implemented elsewhere.

Opsomming

Ontwikkeling van 'n Geïntegreerde Metaboliese-Analise-Gereedskapskis

("Development of an Integrated Metabolic Analysis Toolbox")

C.D. Christensen

Departement Biochemie,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Proefskrif: PhD (Biochem)

Desember 2016

Die lewe is waarskynlik die mees komplekse van alle natuurverskynsels, tog ontstaan dit in wese vanuit dooie molekulêre komponente. 'n Doel van sisteembiologie is om te verstaan hoe eienskappe en nie-lineêre interaksies van hierdie komponente aanleiding gee tot funksies en gedrag van biologiese sisteme. Dit verteenwoordig die sogenaamde "meganistiese verklaring" waarin geen afsonderlike komponent, nog die volledige stelsel self, voorrang geniet.

In hierdie proefskrif word 'n Python-gebaseerde sagteware-pakket voorgestel met die naam `PySCeSToolbox`, wat gepubliseerde teoretiese raamwerke vir die ondersoek van kinetiese modelle van metaboliese sisteme implementeer. Hierdie programmatuur-werktuie is `RateChar`, wat veralgemeende vraag-aanbod analise (VVAA) uitvoer; `SymCa`, wat simboliese metaboliese kontrole-analise uitvoer; en `ThermoKin`, wat onderskei tussen die kinetiese en termodinamiese bydraes tot ensiemgekataliseerde reaksietempo's. Elkeen van die raamwerke soos vervat in die gereedskap van `PySCeSToolbox` beskou metabolisme vanuit 'n ander oogpunt: veralgemeende vraag-aanbod analise gee 'n breë oorsig van die gedrag, beheer, en regulering van metaboliese sisteme met inagneming van hul funksionele organisasie; simboliese kontrole-analise ontleed die beheer-eienskappe van metaboliese stelsels in terme van die fisiese kettings van interaksie tussen ensieme en metaboliese intermediate; en die termodinamiese/kinetiese raamwerk hou die eienskappe van die ensieme self onder 'n vergrootglas om hul regulerende rolle te bepaal. Die krag van `PySCeSToolbox` lê in die integrasie van

hierdie gesigspunte in ’n enkele analise-pakket sodat hulle mekaar kan aanvul in die soektog na ’n meganistiese verklaring van gemodelleerde metaboliese sisteme.

Die toepassing van hierdie sagteware-gereedskapskis in die ondersoek van twee gepubliseerde metaboliese modelle ontbloot nuwe kennis met betrekking tot hul gedrag en verduidelik dit daarna in terme van die eienskappe en interaksies van hul komponente. In ’n model van aspartaat-afgeleide aminosuursintese, wys ’n VVAA dat aspartaat-semialdehied sy produksie-reaksieblok reguleer deur die interaksie met sy vraag-reaksieblokke, ten spyte van die relatief hoë sensitiwiteit van sy produksie-ensiem vir hierdie intermediaat. Hierna word die regulerende bydraes van elk van die vier aspartaat-semialdehied vraag-blokke tot die produksie-reaksieblok gekwantifiseer. In ’n model van die metabolisme van die metaboliese vertakkings rondom pirovaat, toon ’n VVAA dat die fluksie deur ’n $NADH/NAD^+$ vraag-reaksieblok daal wanneer die verhouding van NADH teenoor NAD^+ toeneem. Hierdie verskynsel word ontbloot as ’n gevolg van ’n interaksie van die $NADH/NAD^+$ intermediate met ’n reaksie elders in die pad, en is dus nie ’n gevolg van substraatinhibisie nie.

Simboliese kontrole-analise van die pirovaat-vertakkings-model ontbloot ’n aantal eienskappe wat die nie-ooglopende fluksie-respons, soos bo beskryf, verklaar. Eerstens is slegs enkele beheer-patrone belangrik vir die bepaling van die fluksie-beheer op enige gegewe tydstip. Tweedens domineer verskillende beheer-patrone onder verskillende omstandighede, en die oorheersende patroon verskuif soos hierdie toestande verander. Laastens lei die ontrafeling van hierdie kettings van effekte tot die identifisering van daardie komponente van die sisteem wat verantwoordelik is vir die fluksie-beheer. Bykomende gebruik van die termodinamiese/kinetiese raamwerk om te fokus op daardie ensieme waaruit die beheer-patrone bestaan, herlei hul waardes na die eienskappe van individuele ensiemgekataliseerde reaksies (nl. hul elastisiteite). Hierdie raamwerk word ook gebruik om die waardes van die elastisiteitskoëffisiënt-komponente van die nie-ooglopende fluksie-response te verduidelik, en toon dat hulle hoofsaaklik deur massawerking beheer word. Uiteindelik bied hierdie tweeledige strategie ’n meganistiese verklaring van die fluksie-respons, waarin hierdie hoë-vlak eienskap kwantitatief gekoppel word aan verskeie lae-vlak komponente.

Die ontwerp van `PySCeSToolbox` as ’n Python-gebaseerde biblioteek van programmatuurfunksies vergemaklik die integrasie met die bestaande wetenskaplike Python-ekosisteem, en verskaf dus toegang tot bykomende derde-party sagteware ter ondersteuning van die ontleding van metaboliese sisteme. Hierdie ontwerp moedig ook die gebruik van ’n skripbenadering tot die ontwerp van *in silico* modellerings-eksperimente aan, wat op sy beurt herhaalbaarheid bevorder deur die hergebruik van sodanige skripte. Daarbenewens bied `PySCeSToolbox` rekenaarmatige toegang tot teoretiese analise-raamwerke wat andersins vir navorsers ontoeganklik sou wees omdat dit nêrens anders geïmplementeer is nie.

Acknowledgements

When they say that writing a thesis is difficult, they aren't joking! Many people were instrumental in helping me complete this work, and I would like to extend my greatest thanks to them.

Firstly I would like to thank my supervisor Prof. Johann Rohwer. His way of thinking and working has always inspired me to work harder myself, and his keen eye for detail has kept me on my toes throughout my post-graduate career. The off-track discussions about open source, Python, Linux, and nothing much in particular were always a highlight of our meetings. I would like to especially thank him for his patience and understanding during the times when my health kept me from work, and for his encouragement during the times when work came slowly.

While my co-supervisor Prof. Jannie Hofmeyr only came on board late in my project, he has since given valuable advice for completing my thesis. His way of explaining complicated subject matter in a way that anyone can understand has always been inspiring. I would like to thank him for the time taken to meticulously read through my work and for catching errors that even slipped by Johann.

I also thank my previous co-supervisor Dr. Stéfan van der Walt for the discussions during the early stages of my project and for the time spent together at EuroScipy in 2014. While we did not get to work together much, I truly hope that we will get another opportunity to do so in the future.

I would also like to thank Dr. Johann Eicher and Dr. Danie Palm for their helpful discussions regarding Python programming and Linux, for and the debates about which text editors are the best. `PySCeSToolbox` would have been much more difficult to develop were it not for their help and inspiration.

Additionally, I would like to thank all the members of the Laboratory for Molecular Systems Biology at Stellenbosch University. Were it not for their friendship and the wonderful working environment that they provide (not to mention the work-related discussions), it would have been a struggle to get through these last three and a half years.

I am also truly thankful for the support and encouragement of my family and friends. Having a perpetual student in the family is not easy and I am grateful that I have people in my life that care for me. I especially thank my girlfriend, Leandrie Jacobs. Nobody knows what it took to complete this work better than her, and I am glad that I have her by my side. Without her care, support, and encouragement I might have given up long ago.

Finally, the financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Contents

List of Figures

NOTE

Certain figures in this document to not display correctly when viewed on Mac OSX using the default PDF viewer "Preview". Adobe Acrobat is therefore recommended to ensure that figures are rendered correctly on this platform.

List of Tables

Nomenclature

Abbreviations

ODE	Ordinary differential equation
EFMA	Elementary flux mode analysis
EPA	Extreme pathway analysis
FBA	Flux balance analysis
MCA	Metabolic control analysis
SDA	Supply-demand analysis
GSDA	Generalised supply-demand analysis
PySCeS	The Python simulator for cellular systems
Ac	Acetate
Acal	Acetaldehyde
Acet	Acetoin
Aclac	Acetolactate
Acp	Acetyl phosphate
Glc	Glucose
Lac	Lactate
But	2,3-Butanediol
Pyr	Pyruvate
EtOH	Ethanol
Ado-Met	S-adenosylmethionine
ASA	Aspartate-semialdehyde
Asp	Aspartate
AspP	Aspartyl phosphate
Cys	Cysteine

Hser	Homoserine
Ile	Isoleucine
Lys	Lysine
PHser	Phosphohomoserine
Thr	Threonine
Val	Valine
AK	Aspartate kinase
ASADH	Aspartate-semialdehyde dehydrogenase
HSDH	Homoserine dehydrogenase
DHDPS	Dihydrodipicolinate synthase

Chapter 1

Introduction

Life is arguably the most complex of all natural phenomena; even the most simple organisms consist of an enormous number of interconnected molecular components that are organised into multiple hierarchical functional systems and structures [1, 2]. While the line that separates one functional level or system from another is often unclear [3], it is undeniable that this scheme of organisation plays a central role in the maintenance of living cells and their interaction in the context of the organism. Certainly most biologists would agree that living organisms ultimately amount to collections of molecules, while few would define collections of molecules as living organisms [4]. How is it then that life emerges from dead components?

The study of life at this level, in the form of molecular biology, is a relatively young science when compared to many other biological fields and its inception and development has been greatly influenced by technological advances over the last century. These advances have provided scientists with the means to observe the molecular world with ever greater resolution and fidelity. However, despite the differences between the techniques employed in molecular biology and those of some of its older siblings (such as physiology, botany and zoology), there is still a significant overlap in terms of their approaches and ultimate goals; they are all primarily concerned with characterising and understanding biological entities at some functional level, whether they are whole organisms, organs, or cells. At the molecular level, this entails the identification of the molecules from which biological function arises and the subsequent characterisation of their various properties, behaviours, functions, and interactions with other molecules [5].

To date this approach has been exceptionally successful in characterising life's molecular and chemical components and has even produced intricate maps detailing the relationships between these components within their respective systems. While this wealth of information has had an immeasurable impact on our understanding of the building blocks of life and has

been vital for the advancement of multiple disciplines, such as medicine and agriculture, it has been far less successful in answering our original question regarding life's emergence from these building blocks. It seems that quantitative explanations for how higher level biological function emerges from its molecular components are difficult to derive from descriptions of these components, regardless of their detail and accuracy. Worse still is that this approach has often led to the tendency to attribute higher level function to single components. This situation is perhaps understandable in the light of the vast complexity of these non-linear systems and the inherent limitations of humans in understanding complex patterns.

Systems biology is an approach that aims to overcome these shortcomings. In this paradigm, the concern does not lie with the individual molecular components *per se*, but rather with how their properties and non-linear interactions give rise to higher levels of function [6]. In other words, it seeks to arrive at a system-level understanding of biology. To achieve this goal, systems biology incorporates ideas and techniques from a number of scientific fields outside of biology, such as chemistry, physics, mathematics, and computer science. This multidisciplinary field has to date produced numerous theoretical approaches and methodologies of its own, which have in turn led to many important advancements towards the sought-after system-level understanding.

The study of metabolism is one of the earliest biological fields in which "systems thinking" was applied and form part of what Westerhoff and Palsson [7] call the "systems" root of systems biology. Much of the work in this root has been concerned with computer modelling of metabolic systems and the development new conceptual frameworks for understanding these systems and their models. Metabolic control analysis (MCA) [8, 9] represents a classic example of such a systems biology framework and has been very influential within the field. While MCA was not the first framework used to investigate the sensitivity of metabolic systems [10], this form of metabolic sensitivity analysis has since its development been employed in the study of numerous metabolic systems and has been expanded upon significantly (see [11–13]). MCA has even formed the core of other related frameworks of metabolic analysis [14–17]. Recent years have also seen the increase in scale and complexity of metabolic models, partly due to advances in the "-omics" fields, which constitute the "biology" root of systems biology [7]. One example of a large and complex model is the relatively recently published whole organism model of *Mycoplasma genitalium* by Karr *et al.* [18], which reportedly includes all the molecular components and their interactions together with all annotated gene functions. Models such as these pose new conceptual and technical challenges for systems biologists: How does one investigate a model that approaches the complexity of the system it represents?

Fortunately, many of the frameworks of the "systems" branch have the potential to answer

questions regarding metabolic behaviour in models of any size or complexity. Three such theoretical frameworks are supply-demand analysis [14] and its generalised form [15], symbolic control analysis [19–24] and the related control-pattern analysis [17], and the framework developed Rohwer and Hofmeyr [16, 25] for investigating the kinetic and thermodynamic properties of enzyme-catalysed reactions in metabolic systems. All of the above-mentioned frameworks have, to some degree, been utilised successfully in various metabolic studies [26–31]. Past metabolic studies have, however, not extensively explored the use of these types of techniques to complement one another to answer inter-related questions regarding metabolic function. This lack of complementary application may be due, in part, to the lack of an integrated software implementation of these conceptual tools, and indeed even the lack of software implementations of many of them individually.

The above mentioned frameworks certainly do not represent the only solutions for investigating metabolic behaviour, but in combination they would allow a researcher to follow a thread from the observed systemic behaviour through various levels of description.

1.1 Aims, objectives, and outline

The work presented here is primarily concerned with the development of software that combines existing theoretical techniques into a single metabolic analysis package. While some of these tools are currently available in some form individually, the aim is to develop new implementations for each of the chosen tools. Specifically we set out to develop tools to perform generalised supply-demand analysis, symbolic metabolic control analysis, and to distinguish between kinetic and mass action contributions towards the rate of enzyme-catalysed reactions in metabolic pathways. The second objective is the integration of these tools into a single software package that simplifies and encourages their use in a single work-flow. Finally, we aim to apply these tools in the analysis of metabolic pathways in order to comprehensively quantify their behaviour in terms of the properties and configuration of their components.

Chapter 2, which follows this introduction, reviews the theory behind the conceptual frameworks we wish to implement. In Chapter 3 we develop the software based on the theory laid out in Chapter 2. In Chapter 4 we apply generalised supply-demand analysis to two metabolic models and quantify the contributions of different regulatory routes between intermediate metabolites and their producing and consuming reaction blocks using the `RateChar` package developed as part of the work described in Chapter 3. In Chapter 5 we apply `ThermoKin` and `SymCa`, also developed in the work described in Chapter 3, to one of the models discussed in Chapter 4, thus extending our analysis. Finally, we conclude with a general discussion and critique of the previous chapters in Chapter 6, together with and

examination of possible future work.

Chapter 2

Untangling Metabolic Behaviour: The Tools for Studying Metabolic Systems

While this text is primarily concerned with the development and integrated application of computational tools for investigating metabolism, we must first understand what we mean to investigate. This chapter will therefore commence with an overview of metabolism that briefly highlights some of the features of metabolic systems and how they contribute to metabolic function. Secondly, we will discuss the various methods and frameworks that will be utilised within the metabolic investigations of Chapter 4 and 5. Finally, we will review the various software applications that are currently available for the study of metabolism as a primer for the new software tools presented in Chapter 3.

2.1 Overview of metabolism and its regulation

In the broadest sense, metabolism refers to all the enzyme-catalysed reactions that take place within a living cell. It consists of a well organised, interconnected, open network of enzyme-catalysed reactions that perform the life-sustaining functions of the cell. Here reactions are coupled by their shared chemical intermediates. The complete metabolic network can be subdivided into smaller networks, or metabolic pathways, based on their collective function and the specific intermediates produced by these pathways. Here it is important to note that these subdivisions do not, in reality, represent truly separate entities. Nevertheless, metabolic pathways are effective conceptual frameworks for managing the enormous complexity of metabolism and for understanding its organisation [32].

One of the most important features of metabolic networks is that they have the ability to subsist in a “steady state” where there is a continuous and constant flow of matter, or flux,

through them. At steady state matter enters the metabolic network at the same rate that it exits the network. This results in a zero net change in intermediate concentrations, because they are being produced at the same rate as that at which they are consumed. The steady state is a consequence of the fact that metabolic networks in living cells are open systems; their initial substrates and final products reside outside of the system and the concentrations of these external species are kept constant by the organism through interaction with the environment. At steady state, the external species concentrations are therefore such that they push or pull matter through the metabolic system at a constant rate.

The steady state reflects the functional nature of metabolic systems and can be likened to a factory working at constant capacity to convert raw materials into useful products [14]. While it is not the only state that metabolic networks can reside in (as oscillatory, transient, and chaotic states are also possible [33]) and can in theory only be approached asymptotically, for many metabolic systems it represents a state of homoeostasis that is determined by the evolved properties of the system. The steady state stands in stark contrast to the state of equilibrium found in closed systems. In these self-contained systems there can be no exchange of matter with the environment and they are, therefore, incompatible with life. For these reasons, together with the fact that steady-state models are relatively easy to simulate, the steady state is an important subject within the study of metabolism: many theoretical frameworks, such as those considered within this text, deal exclusively with understanding the steady state, and the end goal for many biotechnological applications is to increase the output of some desirable product by altering the steady state of a system (see, e.g., reviews [34–36]).

The ability to reach a steady state is, however, not unique to metabolic systems as they can arise in any thermodynamically feasible open reaction networks. What sets metabolic networks apart from other open networks is that they have been moulded by evolution to perform specific functions [25, 37]. According to Hofmeyr and Cornish-Bowden [37], this is what is meant when a system is said to be regulated. Metabolic function and metabolic regulation are, therefore, two sides of the same coin [14], and the effectiveness of regulation should be measured in terms of its performance of fulfilling its function. Metabolic function ultimately arises from the design of its network structure, the properties of the enzymes that define the network, and the overall organisation of metabolism.

The structure of a metabolic network is defined by the arrangement of its coupled steps. Most often, these steps are enzyme-catalysed conversions of chemical intermediates, but can also include membrane transport and, much less frequently, uncatalysed reactions. Network reactions are coupled by shared intermediates in a variety of ways, leading to an assortment of structures within the metabolic network: when the product of one reaction acts as the

substrate of the next a linear chain of reactions results (Fig. 2.1A), whereas when two or more reactions share the same substrate or product a branch point occurs (Fig. 2.1B). Additionally, individual steps may also involve the conversion of multiple substrates into multiple products, with the proportions of the changes in mole numbers expressed by the reaction stoichiometry. Combinations of these structures lead to even more complex structures within a metabolic network which ultimately define the network stoichiometry.

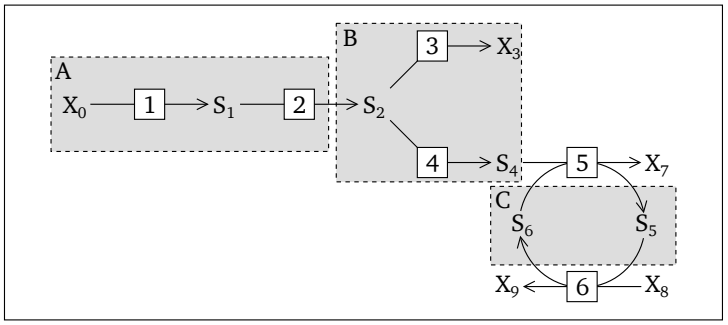


Figure 2.1: A metabolic pathway with different structures. **A** A linear chain of reactions 1 and 2 which are linked by S_1. **B** A branch point due to S_2 acting as a substrate for both reactions 3 and 4. **C** A moiety-conserved cycle of S_5 and S_6 between reactions 5 and 6. Each of reactions 5 and 6 convert two substrates into two products, with each converting one member of the cycle into another.

Network stoichiometry plays an important role in determining the behaviour and functionality of metabolism. It defines the path through which matter can flow, but it also subjects flux to certain constraints. In moiety conserved cycles (Fig. 2.1C), for instance, reactions are coupled in such a way that they both drive, and limit, each other (see [38]). Another example can be found in branch points on the enzyme level where the fluxes through the resulting branches are dependent on each other in a manner unlike that encountered in intermediate level branching [39]. According to Atkinson [40], while certain stoichiometries, such as the coupling of oxidation and reduction, are obligate, others fall into a category called evolved stoichiometries. These stoichiometries represent just one of a number of possible solutions for the chemical conversion of a certain substrate into a product by a sequence of reactions, and they are not fixed by a specific chemical necessity. The evolved stoichiometry is, therefore, a trade-off between different network designs in order to provide an improved solution to the problem of survival. Work by Meléndez-Hevia *et al.* [41] implies that, at least in some cases where maximisation of flux is the goal, optimisation through evolution favours the simplest solution. Whatever the case may be, clearly different structural designs have distinct effects on how metabolism behaves.

The enzymes that constitute the metabolic network are themselves dynamic and functional units. These proteins are highly specific catalysts, with most only binding a single set of intermediates and only catalysing a single reversible reaction. As with uncatalysed chemical reactions, the rates of enzyme-catalysed reactions depend on the concentrations of their substrates and products, however, here catalysis and enzyme binding augment the reaction rate. Another important aspect of enzymes is their ability to be controlled by interactions with intermediates other than those involved in the catalysed reaction. These "allosteric effectors" are intermediates produced elsewhere in the metabolic network and can either activate or inhibit enzyme activity through binding [42, 43]. Enzymatic reaction rates are, therefore, non-linear functions of the concentrations of these intermediates and can be described using various models, or rate equations, based on the enzyme mechanism (see Cornish-Bowden [44]).

As with the network structure, much of metabolic behaviour can be owed to the above-mentioned properties of enzymes. As previously discussed, enzymes largely define the structure of the metabolic network; thermodynamically feasible uncatalysed reactions take place at such slow rates compared to those of the enzyme-catalysed reactions that they can mostly be disregarded. Enzymes, however, do not simply cause a uniform fold change in flux over that of an uncatalysed reaction, but rather act to augment or resist the existing mass-action trend [14, 25]. They are therefore, in part, responsible for determining the steady state of the metabolic network. Additionally, a change in the concentration of any single enzyme will have a unique effect on the properties of the steady state due to its particular characteristics. In other words, enzymes control the flux and intermediate concentrations of the steady state [8, 9]. Similarly, they also determine the responses to changes in the internal or external intermediates as their characteristics define how such a change would propagate throughout the network. Enzymes therefore have multiple levels of function in addition to simple catalysis, and represent one of the main modes through which evolution can shape the functions of metabolic pathways [14].

A final aspect that defines metabolic function which we will touch on, is the organisation of the metabolic network. As reflected by many biochemistry textbooks, the idea that metabolic pathways can be classified according to their function is not new. Typically metabolic processes are simply catagorised as either catabolic or anabolic. Other descriptions are more nuanced and distinguish between degradative metabolism, which provides reducing and phosphorylation power together with carbon skeletons, biosynthesis, which produces molecular building blocks, and macromolecular synthesis and growth, which builds and maintains the structure and machinery of the cell [40]. The common theme between these descriptions is that there is a fanning in from many complex molecules to a few sim-

ple molecules, and a subsequent fanning out from these simple molecules to another set of complex molecules; the functional processes of metabolism are linked by a small subset of common intermediates. Analysis of genome based large-scale metabolic networks has confirmed the validity of these descriptions, which have thus been called the bow-tie structure of metabolism [45, 46].

This scheme of organisation has important implications for the functioning of metabolism. A very obvious issue that is mitigated by the bow-tie structure, is the need for a unique path from each nutrient to each final end product [47]. Therefore this scheme of organisation provides a much simpler solution for dealing with the complex requirements of life and provides robustness in the face of fluctuating environmental conditions; if one nutrient-product path is unavailable, another could be followed. The bow-tie structure is also common in technological systems, such as the power grid, the internet, and manufacturing processes, for the same reason of reducing complexity and increasing robustness [47]. While limiting the number of key metabolites link the producing processes with the consuming processes imparts robustness, it also imparts a type of fragility: disruption of these metabolites could cause a large scale failure [47]. One solution to this problem is functional differentiation, which is the division of labour between the producing and consuming reactions in such a way that homoeostasis of the metabolite is maintained for a wide range of flux variations [14].

Though the metabolic system consists out of numerous interconnected processes, we can clearly see that its design in terms of its overall organisation, its network structure and its functional components facilitates its functionality. While it is not always clear what the function of a metabolic system is and how exactly it achieves this function, it is clear that whatever the purpose of the system is, it stems from its effective regulation and precise design. In the following sections we will concentrate on tools that will help us understand some of the aspects of metabolism described above.

2.2 Modelling metabolic systems

In the light of our overview of metabolism it is clear that its vast complexity precludes it from an intuitive understanding; it is simply impossible for the human mind to keep track of such a huge number of components and their non-linear interactions. Indeed, according to Braess's paradox [48, 49], even networks with few linear interactions can exhibit non-intuitive behaviour. This problem is only exacerbated by the ever increasing rate at which new biological data is being gathered, which is, in part, being facilitated by high throughput technologies that are able to simultaneously keep track of thousands of cellular components. Thus with limited brain power and seemingly unlimited biological data, the only avenue

for reaching an understanding of metabolism (and biological systems in general) is via the modelling and simulation of these systems.

A model is a simplified representation of a system or process which aims to capture its essential properties. Many models include only those components of the system necessary to accurately portray the most important aspects of the system in question. In recent times, however, there has been a shift towards the construction of large genome scale models (e.g. [18, 50–52]), which include information about hundreds or thousands of cellular components. In either case, models are constructed by translating the properties of the system and its components into mathematical language. Ultimately the goal of models is to allow for an understanding of how the system's underlying structure and components lead to its functions and behaviour.

To date a number of different types of models and approaches have been developed, each with its own set of outcomes and limitations. The most simple type of metabolic models are arguably structural models (see e.g. [53]), which only capture information about the stoichiometry and topology of the system in question. Kinetic models, on the other hand, include another level of detail in the form a description of the kinetics of the enzyme-catalysed reactions in the system (e.g. [34, 54, 55]). Kinetic models are the most relevant for the purpose of this dissertation, however, as there is a great overlap in how both types of models are defined we shall begin our discussion of the modelling procedure with a description of structural models.

2.2.1 Structural models

As previously discussed, metabolic pathways consists of numerous coupled reactions where the proportions of molecules consumed and produced in a single reaction is expressed by the reaction stoichiometry. In reaction 5 of the metabolic pathway shown in Fig. 2.1, for example, the stoichiometric coefficients are -1, -1, 1, and 1, for S_4, S_6, S_5, and X_7 respectively. The negative coefficients indicate the substrates of the reaction while the positive coefficients indicate the products. Using these principles, the stoichiometry of a complete reaction network can be expressed as a stoichiometric matrix $\mathbf{N}$ (see e.g [13, 56, 57]). Thus the stoichiometric

matrix for the pathway in Fig. 2.1 is:

$$\mathbf{N} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_4 \\ S_5 \\ S_6 \end{array} \begin{array}{c} \begin{array}{cccccc} R1 & R2 & R3 & R4 & R5 & R6 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix} \end{array} \tag{2.1}$$

In this $m \times n$-dimensional matrix, the rows represent the species while the columns represent the reactions of the pathway. Each element c_{ij} of $\mathbf{N}$ therefore represents the stoichiometric coefficient of the species S_i participating in the reaction j [57]. Note that the external species (e.g. X_7) are not included in the stoichiometric matrix, as their concentrations are considered to remain constant.

Using the stoichiometric matrix we can express the time-dependent change in intermediate concentrations of a system as a set of ordinary differential equations (ODEs):

$$\frac{d\mathbf{s}}{dt} = \mathbf{N}\mathbf{v}, \tag{2.2}$$

where $\mathbf{s}$ is a m-dimensional column vector of metabolite concentrations and $\mathbf{v}$ is a n-dimensional column vector of reaction rates. As previously discussed the steady state is characterised by a zero net change in intermediate concentrations, therefore Equation 2.2 is equal to $\mathbf{0}$ in this state. This forms the basis of structural models. While kinetic models are also based on Equation 2.2, its value need not be zero for these models, as they are not required to subsist in a steady state for sensible analysis.

2.2.1.1 The kernel matrix

As previously mentioned, structural models do not contain information about reaction kinetics. In other words, instead of $\mathbf{v}$ being treated as set of functions (as will be discussed in the next section), $\mathbf{v}$ is treated as a set of variables. It is, however, possible to determine how flux is distributed in a metabolic system using only this structural information; as the value of $\mathbf{N}$ is known and $\mathbf{N}\mathbf{v} = \mathbf{0}$, the null space of the stoichiometric matrix contains all the solution vectors of $\mathbf{v}$. We can express this as:

$$\mathbf{N}\mathbf{K} = \mathbf{0}, \tag{2.3}$$

where the columns of the kernel matrix $\mathbf{K}$ represent linearly independent solution vectors, or flux modes, satisfying the steady-state condition. With the fluxes denoted with J, and fluxes

J_3 and J_6 arbitrarily chosen as independent fluxes, the flux relationships ($\mathbf{J} = \mathbf{K}\mathbf{J_i}$) for the pathway in Fig. 2.1 can be expressed with the $\mathbf{K}$ matrix as [57]:

$$\underbrace{\begin{bmatrix} J_3 \\ J_6 \\ J_1 \\ J_2 \\ J_4 \\ J_5 \end{bmatrix}}_{\mathbf{J}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} J_3 \\ J_6 \end{bmatrix}}_{\mathbf{J_i}} \tag{2.4}$$

where the column vector $\mathbf{J}$ contains all the fluxes in the system, and the column vector $\mathbf{J_i}$ contains the independent fluxes. All the possible flux distributions of the network can thus be determined through linear combinations of the flux modes contained in columns of $\mathbf{K}$ [56, 58].

2.2.1.2 Elementary flux modes and extreme pathways

The disadvantage of using the kernel matrix is that it provides a non-unique description of the flux modes, i.e., the solutions contained in the columns can themselves be expressed in terms of simpler non-decomposable flux modes. A solution to this problem can be found in the two related techniques of elementary flux mode analysis (EFMA) [58–60] and extreme pathways (EPA) [50, 61]. In essence, both techniques take into account the reversibility of the reactions in a pathway and involve the application of an inequality constraint on the flux values of the irreversible reactions of the pathway. In both techniques a unique set of flux modes are generated for any given pathway. Additionally, no reaction can be removed from any flux mode if it is to retain its functionality.

The difference in these two techniques lies in how they treat reversible and irreversible reactions. In EFMA, a set of rules account for the directionality of the reactions during the calculation of the flux modes, whereas in EPA reversible reactions are treated as two separate reactions operating in opposite directions [53]. This difference in treatment results in a smaller number of extreme pathways than elementary modes for any given pathway. Here the extreme pathways are a subset of elementary modes that cannot be represented by a non-negative linear combination of any other extreme pathways. In the case where all the reactions involving external metabolites (exchange reactions) are irreversible, however, EFMA and EPA will yield the same flux modes [62].

2.2.1.3 Flux balance analysis

Another technique that can be applied to structural models is flux balance analysis (FBA) [63–65]. Instead of characterising the complete solution space of flux distributions, FBA is used to search for those flux distributions which are optimal relative to some criteria [65]. To this end, $\mathbf{Nv} = \mathbf{0}$ is cast as a linear programming problem. Here reaction rates are given upper and lower bounds which conform to experimentally measured fluxes, or to maximal or minimal rates. Additional constraints may also be included based on specific experimental conditions or assumptions. Most importantly, an objective function is supplied to which the problem must be solved under the steady- state and reaction rate constraints.

The choice of the objective function is based on the problem being studied. A typical example of an objective function is the maximisation of growth rate. The choice of the objective function will yield a flux distribution which, while conforming to the constraints, differentiates between those reactions which are essential for achieving this objective, and which ones are not [65]. Different objective functions will typically yield different solutions, however multiple solutions may exist for the same objective function [66, 67].

Flux balance analysis can be a powerful tool for studying metabolism. In *E. coli*, for instance it has been shown that FBA could predict the effect on growth rate by various gene deletions with an accuracy of 86% [68]. Similarly, it has been shown that certain objective functions and constraints are more accurate than others in predicting flux distributions under a variety of conditions in the same organism [69]. Another use for FBA is the prediction of yields for cofactors such as ATP and NADH [70]. As with EMFA and EPA, FBA is also useful for analysing large genome-scale models where kinetic data is unavailable or difficult to incorporate [71, 72]. One group of authors have suggested that FBA could be used to answer more "profound" questions such as "why microorganisms are not optimally efficient in energetic terms" [73]. A common criticism of FBA, however, is that results yielded do not necessarily accurately depict the flux distributions found in real systems, in part due to the difficulty of choosing biologically relevant objective functions [55, 73]. Various approaches to solving this problem have however been put forward in the form of algorithms for automatically selecting biologically plausible objective functions [74–76].

2.2.2 Kinetic models

As with structural models, kinetic models are based on the set of ODEs of Equation 2.2, however, they are differentiated from their counterparts by their inclusion of enzyme kinetics. Thus, in these models $\mathbf{v}$ is treated as a function:

$$\mathbf{v} = \mathbf{v}[\mathbf{s}, \mathbf{p}], \tag{2.5}$$

where **s** is the familiar metabolite concentration vector and **p** is a p-dimensional column vector of parameters [57]. These models have the potential to accurately predict the properties of the steady state and can be used to provide mechanistic explanations for system behaviour. However, the increase in fidelity that these models bring over structural models is at the cost of being more cumbersome to construct, due to the necessary inclusion of enzyme kinetic data for each reaction within the modelled system.

2.2.2.1 Enzyme kinetics

Enzyme kinetics is the study of enzyme-catalysed reaction rates. As previously described in Section 2.1, and implied by Equation 2.5, the rates of these reactions can be described in terms of the concentrations of the species involved in the reaction as well as various parameters. These reaction rates can be described using a function, or rate law, which describes their dependence on **s** and **p**.

The most well known example of a rate law is probably the Michaelis–Menten equation. This rate law was first published in 1913 as description of the reaction rate of the enzyme invertase using the concept of an enzyme-substrate complex [77]. It describes the unidirectional conversion of a single substrate, S into a single product P and can be expressed as:

$$v = \frac{V_{max} s}{K_m + s} \tag{2.6}$$

where V_{max} is the limiting rate, K_m is the Michaelis constant and s is the concentration of S. At low substrate concentrations this equation can be simplified to $v = V_{max}/K_m \times s$, where the reaction rate responds linearly to s. At higher concentrations of S, however, the reaction rate responds increasingly less linearly to further increases in s, and eventually reaches a point where it asymptomatically tends to a maximum rate defined by V_{max}. Here V_{max} is a function of the total enzyme concentration, therefore this phenomenon occurs due to the saturation of the all available free enzyme. The Michaelis–Menten equation has been vital for understanding the functioning of enzymes and its creators are considered to be, in part, responsible for the inception of the field of enzymology [44].

Reactions within metabolism are, however, very often more complicated than the monomolecular irreversible reaction described by the Michaelis–Menten equation. Right out of the gate this equation is precluded from describing any of the numerous reversible and multimolecular reactions found in metabolism. Among these multi-substrate enzymes there also exists a variety of different mechanisms governing the sequence and method of binding. Furthermore, in the case where multiple substrate molecules bind an enzyme (such as in the case of multimeric enzymes), the binding of one substrate molecule can affect the binding

of the next in a process known as cooperativity. Moreover, as previously discussed, enzymes can often interact with allosteric effectors which modify their rates [42, 43]. In order to describe this large variety of enzyme mechanisms a number of different rate equations have been developed: The Michaelis–Menten equation has been generalised for reversible multimolecular reactions [44] and the Adair [78], Hill [79, 80], and Monod-Wyman-Changeux models [43, 81], for example, describe enzymes exhibiting cooperative binding. Rate equations accounting for binding order (such as ternary-complex and substituted-enzyme mechanism) have also been developed [44].

These rate equations were, as implied, originally developed as a means to understand enzyme mechanisms, rather than to characterise the rate of enzymes at different concentrations or conditions. In recent times, however, there has been a shift in focus towards simpler, generalised rate equations that do not necessarily accurately describe the enzyme mechanism [49, 55, 82, 83]. This is due to a variety of reasons, one of which is that the full mechanistic characterisation of some enzymes requires a large amount of data [80]. Even if only the forward reaction is characterised, it may not be an adequate description for use in metabolic models. Enzyme characterization is thus a cumbersome process, especially if the main purpose of the characterisation is not to understand the enzyme mechanism, but to describe the rate of a reaction. Another important reason for the shift towards simpler rate equations is that, as long as they can closely approximate the behaviour of an enzyme, they can often be used in metabolic models without affecting their accuracy or predictive power. This is demonstrated in a variety of successful models which either make simplifying assumptions or explicitly use simplified rate equations (e.g. [84–87])

An example of such a simplified rate equation is the reversible Hill equation [80]. This generalisation of the Hill equation was developed in order to account for reversible reactions. For a the reaction:

$$\mathrm{S} \rightleftharpoons \mathrm{P} \tag{2.7}$$

with a single allosteric modifier M that either inhibits or activates the enzyme, the rate equation is expressed as :

$$v = \frac{V_f \sigma \left(1 - \frac{\Gamma}{K_{eq}}\right)(\sigma + \pi)^{h-1}}{\frac{1+\xi^h}{1+\alpha^{2h}\xi^h} + (\sigma + \pi)^h} \tag{2.8}$$

where σ, π, and ξ are respectively the concentrations of S, P and M scaled by their half-saturation constants $S_{0.5}$, $P_{0.5}$ and $M_{0.5}$. Here these half-saturation constants are equivalent to the Michaelis constant in Equation 2.6. Furthermore Γ is the mass-action ratio p/s and K_{eq} is the equilibrium constant. Finally h is the Hill coefficient, which is used to quantify the degree of cooperativity (with values typically between 1 and 4 for positive cooperativity),

and α is an interaction factor used to quantify the effect of modifier binding on substrate and product binding (with $\alpha > 1$ indicating and activation $\alpha < 1$ indicating inhibition).

In order to address the fact that the reversible Hill equation as presented in [80] can only account for monomolecular reactions and at most two allosteric effectors a generalised rate equation specifically for systems biology was developed by [83]. Using the Hill equation as a base, the "universal rate equation" is generalised for both an arbitrary number of substrate- product pairs, and an arbitrary number of either independent or competing allosteric effectors. Moreover this rate equation addresses the unwieldiness of mechanistically characterising complex reactions as it requires far fewer experimental measurements, while still providing results that adequately describe the reaction rate for the purposes of metabolic modelling.

In Section 2.6 we will continue our discussion of enzyme kinetics by exploring how to distinguish between the thermodynamic and kinetic effects on reaction rate.

2.2.2.2 Time-course simulations and steady-state analysis

The kinetic model defined in Equation 2.2 is, as shown above, a set ordinary differential equations. Due to their non-linearity, finding analytical solutions to these sets of equations is infeasible; however, they can be solved in a number of ways using methods commonly available in general purpose computer mathematics packages [49] such as SageMath [88], Matlab [89] or Mathematica [90]. Typically, however, kinetic models are simulated using one of the numerous software packages developed specifically for biological models as will be reviewed in Section 2.7.

Regardless of the software used for metabolic modelling, two types of analyses are most common [49, 55]. The first is time-course simulations, where, starting from an initial set of metabolite concentrations, the evolution of the reaction rates and metabolite concentrations over time is determined. When starting from a set of arbitrary initial conditions, the values of the metabolic variables will typically increase or decrease over time, with each eventually reaching a point where no further change takes place. Once each variable has become stable, the system has reached a steady state. Time-course simulations are an example of an initial value problem and are commonly solved (e.g. [91, 92]) using the LSODA solver [93, 94].

The second type of common analysis is steady-state analysis. Here Equation 2.2 is set to $\mathbf{0}$ and the set of ODEs is solved. Thus, the values of $\mathbf{v}$ and $\mathbf{s}$ are determined, and should be equivalent to the steady-state values produced by a time-course simulation. Here the NLEQ2 solver [95] can used to determine the steady-state solution (e.g. [91]).

Another type of analysis that can be performed on kinetic models is metabolic control analysis [8, 9]. While the expressions of the control coefficients can be determined through

the application of the summation and connectivity theorems, a kinetic model at steady state is required in order to determine the values of the coefficients of MCA. The next two sections will be dedicated to the discussion of this method.

2.3 Metabolic control analysis

Metabolic control analysis was originally developed independently in the early 1970s by Kacser and Burns [8], and by Heinrich and Rapoport [9] in order to establish a general theory on the control of metabolic systems. This type of sensitivity analysis quantifies the global control properties of a system in terms of the responses of its fluxes and steady-state metabolite concentrations towards perturbations in the rates of its reactions. Furthermore, it relates these responses to the local responses of the system's enzymes towards perturbations in their affecting parameters. Clearly MCA conforms to the "systems biology paradigm" in the sense that it relates higher level systemic functionality to the properties of, and interactions between, lower level components [55].

This framework provides us with the tools and language [96] to describe the effects of the various system components on the properties of the system and forms the theoretical basis of much of the work described in this dissertation. Since its inception it has been expanded upon by numerous researchers spanning hundreds of publications, much of which is collected and summarised in two comprehensive books, respectively authored by Fell [12], and by Heinrich and Schuster [13]. Interestingly, both original groups of authors [8, 9] refer to the difficulty of gaining deeper insight into the control properties of a metabolic system based purely on computer simulations as a reason for the development of MCA, thus drawing a strong parallel with the motivations behind the current work.

Central to MCA is its use of three ratios of change, or "coefficients", to describe metabolic control; each being classified as either global (referring to systemic properties) or local (referring to component properties). Below we will discuss these coefficients using the standard nomenclature defined in 1985 [96], together with the summation and connectivity theorems [8].

2.3.1 Elasticity coefficients

Elasticity coefficients [96], or simply elasticities, describe the sensitivities of the functional components of metabolic pathways towards infinitesimal changes in the variables or parameters that affect them. Here "functional components" usually refer to the enzymes of a system, but transporters and non-enzymatic reactions can also be described using elasticity coefficients. In the case of an enzyme its affecting variables include its substrates, products, acti-

vators and inhibitors, while its parameters are the thermodynamic properties inherent to the reaction being catalysed, such as the equilibrium constant (K_{eq}), or the kinetic properties of the enzyme itself, such as the binding (K_M) or catalytic constants (k_{cat}). Invariant properties of system, such as the concentrations of external metabolites or the total concentration of a conserved moiety, are also regarded as parameters.

Elasticity coefficients are typically expressed as a fractional change in a reaction rate towards a fractional change to a variable or parameter. The elasticity coefficient of reaction step i with rate v_i towards a variable or parameter x will be defined as:

$$\varepsilon_x^{v_i} = \frac{\partial v_i / v_i}{\partial x / x} \tag{2.9}$$

Equivalently, elasticity coefficients may also be expressed as the partial derivative of $\ln v_i$ with respect to $\ln x$:

$$\varepsilon_x^{v_i} = \frac{\partial \ln v_i}{\partial \ln x} \tag{2.10}$$

This second form is very useful, because it allows us to read off the elasticity coefficient as a slope of the tangent of a plot in double-logarithmic space of reaction rate against variable or parameter value. This advantage also applies to the two remaining coefficients which may likewise be read off from their appropriate plots. For this reason we will favour the use the second form of the definition in this dissertation (see Section 2.5.1).

Elasticity coefficients are classified as local properties, because they describe the sensitivity of a system component in isolation; all variables and parameters other than the one of interest remain constant. In simple terms the elasticity coefficient of an enzyme will not be affected by any other component in a system, and single set of conditions will produce the same elasticity coefficients regardless of the system the enzyme finds itself in.

2.3.2 Control coefficients

Control coefficients [96] describe the sensitivity of a steady-state flux or metabolite concentration towards infinitesimal changes in the activity of reaction steps in the system. Similar to elasticity coefficients, they can be defined as fractional changes or as derivatives in logarithmic space. The sensitivity of a steady-state variable y towards a change in the rate v_i of step i will be defined as:

$$C_i^y = \frac{dy/y}{dv_i/v_i} = \frac{d\ln y}{d\ln v_i} \tag{2.11}$$

Unlike elasticity coefficients, control coefficients are global or systemic properties. Clearly this must be the case as they describe the sensitivities of steady-state variables, which are systemic properties themselves. Once more, this means that they describe the properties arising

from the stoichiometry and the properties of the components of a systems. Any structural or kinetic alterations will therefore propagate through the system, leading to new control coefficients.

2.3.3 Response coefficients

Response coefficients [8] are also global properties that describe the sensitivity of a steady-state variable towards an infinitesimal change in a system parameter and are defined similarly to the two previously discussed coefficients. The response coefficient of the steady-state variable y towards an infinitesimal change in the parameter x which affects reaction step i in the system is defined as:

$$^{i}R_{x}^{y} = \frac{dy/y}{dx/x} = \frac{d\ln y}{d\ln x} \tag{2.12}$$

In the case where parameter x affects y through multiple reaction steps the response coefficient is the sum of all the response coefficients describing these individual sensitivities:

$$R_{x}^{y} = \sum_{i=1}^{n} {}^{i}R_{x}^{y} \tag{2.13}$$

where n is the total number of steps in the system. The individual right-hand terms above (as defined in equation 2.12) are called partial response coefficients.

It is important to note that response coefficients can only be non-zero for system parameters. This is because internal variable perturbations cause non-zero *transient responses* in a system which allows it to converge towards its original state and counteract the initial perturbation; no measurable difference between the two states is produced. This does not necessarily apply to systems that undergo metabolic oscillations or that have multiple steady states, but it is true for systems with a single steady state such as those considered in this dissertation.

2.3.4 The partitioned response property

As the reader may have deduced, there is a relationship between the three previously defined coefficients. Termed the partitioned or combined response property, it stems from the fact that the local effect of a parameter change on enzyme activity can be related to a system variable via its sensitivity towards the enzyme activity in question. In mathematical terms this can be stated as:

$$\frac{d\ln y}{d\ln x} = \frac{d\ln y}{d\ln v_i} \cdot \frac{\partial \ln v_i}{\partial \ln x}$$

or simply:

$$^{i}R_x^y = C_i^y \varepsilon_x^{v_i} \tag{2.14}$$

This definition allows us to rewrite equation 2.13 as:

$$R_x^y = \sum_{i=1}^{n} C_i^y \varepsilon_x^{v_i} \tag{2.15}$$

which is the generalised form of the partitioned response property that applies when a parameter affects multiple steps in a system.

2.3.5 The summation theorem

The summation theorem [8] is a property that describes the distribution of control of the metabolic variables between the reactions of a pathway. It states that for any metabolic system, the sum of the control coefficients of all the reactions on any particular flux will be 1. Furthermore it does not depend on any specific metabolic features, such as the size, structure, or the reaction properties of a system, and therefore applies to all metabolic systems. For a system of n reactions this property can be expressed as:

$$C_1^J + C_2^J + \cdots + C_n^J = 1 \tag{2.16}$$

or more generally as:

$$\sum_{i=1}^{n} C_i^J = 1 \tag{2.17}$$

Multiple mathematical proofs for this theorem exist (see [12]), but we will rather focus on its meaning within control analysis. It shows that the flux control coefficients of the reactions within any linear metabolic pathway will have values between zero and one as long as those reactions have normal kinetics (meaning that they are not substrate inhibited or product activated). Importantly, this implies that unless all but one of the control coefficients are zero, control of flux will be shared between the reactions.

In branched pathways the situation is somewhat different as the flux control coefficient in one branch may be negative with respect to the reactions in a different branch. Intuitively this makes sense as one would expect an increase in flux in one branch, stemming from an increase in the activity of one of its reactions, to decrease the flux in a competing branch, i.e., exerting negative control.

The summation property also applies to steady-state metabolite concentrations [9], however here the concentration control coefficients of all the reactions with respect to a particular

metabolite will sum to 0. For metabolite S_j, with concentration s_j, in a system with n reactions it is expressed as:

$$C_1^{s_j} + C_2^{s_j} + \cdots + C_n^{s_j} = 0 \tag{2.18}$$

or more generally as:

$$\sum_{i=1}^{n} C_i^{s_j} = 0 \tag{2.19}$$

The zero value of the concentration summation property implies that some reactions will exert negative control on a concentration, while some will exert positive control. This means that unless all control coefficients are zero, there must be at least one positive and one negative control coefficient.

2.3.6 The connectivity theorem

While the summation theorem describes the relationship between different reactions in terms of their control with respect to a particular metabolic variable, the connectivity theorem [8] relates control coefficients to elasticity coefficients. This theorem states that in a system where multiple reactions are affected by a particular internal metabolite the sum of the products of the flux control coefficients with respect to these reactions and their corresponding elasticity coefficients with respect to the internal metabolite will be zero. In a system where S_j, with concentration s_j, affects reactions 1, 2, and 3, this property can therefore be expressed as:

$$C_1^J \varepsilon_{s_j}^{v_1} + C_2^J \varepsilon_{s_j}^{v_2} + C_3^J \varepsilon_{s_j}^{v_3} = 0 \tag{2.20}$$

This equation can also be expressed in the general form:

$$\sum_{i=1}^{n} C_i^J \varepsilon_{s_j}^{v_i} = 0 \tag{2.21}$$

where n is the number of reactions in the pathway. While this form includes even reactions not affected by S_j, the property still holds as the elasticity coefficients corresponding to those reactions will be zero. The physical interpretation of this equation is that a system returns to its initial state after an intermediate perturbation, thus the net effect of such a perturbation is zero.

As with the summation theorem, there is a version of the connectivity theorem that applies to steady-state metabolite concentrations [97]. Here the right-hand term depends on which metabolites are being considered. In a system where the concentration control of the metabolite S_k, with concentration s_k, is considered together with the elasticities of the reactions affected by the concentration s_j of metabolite S_j, we can arrive at the general expression:

$$\sum_{i=1}^{n} C_i^{s_k} \varepsilon_{s_j}^{v_i} = -\delta_{jk} \tag{2.22}$$

In the case where $j = k$ the Kronecker delta $\delta_{jk} = 1$, while if $j \neq k$ then $\delta_{jk} = 0$.

2.4 Symbolic control analysis

The summation and connectivity theorems expressed above are, to a great extent, some of the most powerful features of MCA in that they respectively demonstrate the distribution of control in metabolic systems, and relate the properties of the enzymes to the control properties of the system. The latter feature can be taken a step further by solving these simultaneous equations to produce expressions for each control coefficient in terms of elasticity coefficients. Thus for a simple three step pathway:

$$X_0 \xrightarrow{1} S_1 \xrightarrow{2} S_2 \xrightarrow{3} X_3 \tag{2.23}$$

with the summation property of:

$$C_1^J + C_2^J + C_3^J = 1 \tag{2.24}$$

and the connectivity properties of:

$$C_1^J \varepsilon_{s_1}^{v_1} + C_2^J \varepsilon_{s_1}^{v_2} = 0, \tag{2.25}$$

and

$$C_2^J \varepsilon_{s_2}^{v_2} + C_3^J \varepsilon_{s_2}^{v_3} = 0 \tag{2.26}$$

we can derive expressions for C_1^J, C_2^J, and C_3^J, as follows:

Collect C_1^J in Equation 2.25 and C_3^J in Equation 2.26

$$C_1^J = -\frac{C_2^J \varepsilon_{s_1}^{v_2}}{\varepsilon_{s_1}^{v_1}} \tag{2.27}$$

$$C_3^J = -\frac{C_2^J \varepsilon_{s_2}^{v_2}}{\varepsilon_{s_2}^{v_3}} \tag{2.28}$$

Substitute the right-hand sides of Equations 2.27 and 2.28 into Equation 2.24

$$-\frac{C_2^J \varepsilon_{s_1}^{v_2}}{\varepsilon_{s_1}^{v_1}} + C_2^J + -\frac{C_2^J \varepsilon_{s_2}^{v_2}}{\varepsilon_{s_2}^{v_3}} = 1 \tag{2.29}$$

$$\therefore \quad C_2^J = \frac{-\varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_3}}{\varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_2} + \varepsilon_{s_1}^{v_2}\varepsilon_{s_2}^{v_3} - \varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_3}} \tag{2.30}$$

Substitute the right-hand side of Equation 2.30 into Equations 2.25 and 2.26

$$C_1^J = \frac{\varepsilon_{s_1}^{v_2}\varepsilon_{s_2}^{v_3}}{\Sigma} \tag{2.31}$$

$$C_3^J = \frac{\varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_2}}{\Sigma} \tag{2.32}$$

$$\text{where} \quad \Sigma = \varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_2} + \varepsilon_{s_1}^{v_2}\varepsilon_{s_2}^{v_3} - \varepsilon_{s_1}^{v_1}\varepsilon_{s_2}^{v_3} \tag{2.33}$$

At this point it should be clear that, in spite of the simplicity of the 3-step pathway above, deriving the control coefficient expressions can be quite cumbersome. In systems of any realistic complexity, this problem is magnified and these derivations become an enormous undertaking.

2.4.1 The control-matrix equation

Fortunately a number of different matrix based formalisations of MCA have been developed (e.g. [19–21, 56, 57, 98–110]) in which the derivation of the control coefficients from the elasticity coefficients is greatly simplified. This is made possible through what is known as the control-matrix equation, which combines the connectivity and summation theorems into a single expression. The original incarnation of control-matrix equation was first derived by Reder [56] in 1988, who demonstrated that this relationship can be derived in steady-state systems of arbitrary complexity, and that it is a natural consequence of the structural features of these systems. This work was an important development in the field of MCA and it addressed, in part, some of the criticisms [111–113] against its the general applicability [21]. Since the inception of the control-matrix equation, much work has gone into refining this method by a number of different groups (e.g [21, 107, 109, 110]). The remainder of this text, however, will refer to the form developed by Hofmeyr and Cornish-Bowden [107, 110], specifically as laid out in [57]. Unlike the approach by Reder [56], this form uses scaled coefficients rather than unscaled coefficients. The benefit here is that scaled coefficients, while more unwieldy in a mathematical sense, are more biologically meaningful than their unscaled counterparts as noted by various authors (e.g. [13, 25]). In order to express the control-matrix equation we will first build on some of the concepts established in Section 2.2.1.

The first concept we will introduce is that of link matrix. If r is the rank of $\mathbf{N}$ and $r = m$ (where m is the number of variable species, see Section 2.2.1), then all the rows of $\mathbf{N}$ are

linearly independent. This means that there are no conservation relationships between the intermediates of the system represented by **N**. In the case where $r < m$, however, then there are $m-r$ dependencies among the differential equations, thus indicating that only r rows are linearly independent. This means there are r independent metabolites and $m-r$ dependent metabolites. In this case it is possible to express the time derivatives of all of the intermediates in terms of the time derivatives of the independent intermediates using the $m \times r$-dimensional link matrix **L**:

$$\frac{d\mathbf{s}}{dt} = \mathbf{L}\frac{d\mathbf{s_i}}{dt} \tag{2.34}$$

where $\mathbf{s_i}$ is an r-dimensional vector of independent metabolite concentrations. Additionally, the matrix **L** has the structure:

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{L_0} \end{bmatrix} \tag{2.35}$$

where $\mathbf{I}_r$ is an r-dimensional identity matrix and $\mathbf{L_0}$ is an $(m-r) \times r$-dimensional matrix expressing the time derivatives of the dependent intermediates in terms of the time derivatives of the independent intermediates.

Next, as demonstrated in Equation 2.4, the kernel matrix **K** can be used to express the relationships between the dependent and independent steady-state fluxes. **K** is an $n \times (n-r)$-dimensional matrix which links the n-dimensional column vector of all the steady-state fluxes **J** with the $n-r$-dimensional vector of independent steady-state fluxes $\mathbf{J_i}$. Internally, **K** has the structure:

$$\mathbf{K} = \begin{bmatrix} \mathbf{I}_{n-r} \\ \mathbf{K_0} \end{bmatrix} \tag{2.36}$$

where $\mathbf{I}_{n-r}$ is an $n-r$-dimensional identity matrix and $\mathbf{K_0}$ is an $r \times (n-r)$-dimensional matrix expressing the dependent fluxes in terms of the independent fluxes.

It is important to note that the arrangement of the rows of **L** and **K** as given respectively in Equations 2.35 and 2.36 corresponds to the arrangements of the column vectors **J** and **s** respectively. In **J** and **s**, the rows have been arranged in such a way that the independent variables are stacked on top of the dependent variables. In each of the following matrices that will be introduced, the rows will be ordered in the same manner.

In order to construct the control-matrix equation, the **L** and **K** matrices need to be scaled. For this purpose the diagonal matrices $\mathbf{D^J}$ and $\mathbf{D^s}$ are defined, which respectively have fluxes and steady-state concentrations on their diagonals. The inverses of these two matrices, $(\mathbf{D^J})^{-1}$ and $(\mathbf{D^s})^{-1}$, therefore respectively have inverses of fluxes and inverses of steady-state concentrations on their diagonals. Additionally, the diagonal matrices $\mathbf{D^{J_i}}$ and $\mathbf{D^{s_i}}$ are similarly defined for independent fluxes and independent steady-state concentrations. Thus **L** is nor-

malised to produce $\mathcal{L}$:

$$\mathcal{L} = (\mathbf{D}^{\mathbf{s}})^{-1}\mathbf{L}\mathbf{D}^{\mathbf{s_i}}, \tag{2.37}$$

whereas $\mathbf{K}$ is normalised to produce $\mathcal{K}$:

$$\mathcal{K} = (\mathbf{D}^{\mathbf{J}})^{-1}\mathbf{K}\mathbf{D}^{\mathbf{J_i}} \tag{2.38}$$

Now the control-matrix equation for the control coefficients of the independent metabolic variables $\mathbf{C}^{\mathbf{i}}\mathbf{E} = \mathbf{I}_n$ can be constructed [57]:

$$\begin{bmatrix} \mathbf{C}^{\mathbf{J_i}} \\ \mathbf{C}^{\mathbf{s_i}} \end{bmatrix} \left[\mathcal{K} - \boldsymbol{\varepsilon}_{\mathbf{s}}\mathcal{L}\right] = \mathbf{I}_n \tag{2.39}$$

where
$\mathbf{C}^{\mathbf{J_i}}$ is an $(n-r)\times n$-dimensional matrix of independent flux control coefficients,
$\mathbf{C}^{\mathbf{s_i}}$ is an $r \times n$-dimensional matrix of independent concentration control coefficients,
$\boldsymbol{\varepsilon}_{\mathbf{s}}$ is an $n \times m$-dimensional species elasticity matrix, and
$\mathbf{I}_n$ is an n-dimensional identity matrix.
Here both $\mathbf{C}^{\mathbf{i}}$ and $\mathbf{E}$ are invertible $n \times n$-dimensional square matrices. Thus it follows that by inversion of $\mathbf{E}$ we can determine $\mathbf{C}^{\mathbf{i}}$, i.e.:

$$\mathbf{C}^{\mathbf{i}} = \mathbf{E}^{-1} \tag{2.40}$$

Finally, the remaining control coefficients for the dependent metabolic variables contained in $\mathbf{C}^{\mathbf{s_d}}$ and $\mathbf{C}^{\mathbf{J_d}}$ can respectively be obtained from the relationships:

$$\mathbf{C}^{\mathbf{s_d}} = \mathcal{L}_{\mathbf{0}}\mathbf{C}^{\mathbf{s_i}}, \tag{2.41}$$

and

$$\mathbf{C}^{\mathbf{J_d}} = \mathcal{K}_{\mathbf{0}}\mathbf{C}^{\mathbf{J_i}}, \tag{2.42}$$

where $\mathcal{L}_{\mathbf{0}}$ and $\mathcal{K}_{\mathbf{0}}$ are respectively the scaled forms of $\mathbf{L}_{\mathbf{0}}$ and $\mathbf{K}_{\mathbf{0}}$ introduced in Equations 2.35 and 2.36. These relationships are contained in the full control-matrix equation:

$$\begin{bmatrix} \mathbf{C}^{\mathbf{J_i}} \\ \mathbf{C}^{\mathbf{J_d}} \\ \mathbf{C}^{\mathbf{s_i}} \\ \mathbf{C}^{\mathbf{s_d}} \end{bmatrix} \left[\mathcal{K} - \boldsymbol{\varepsilon}_{\mathbf{s}}\mathcal{L}\right] = \begin{bmatrix} \mathbf{I}_{\mathbf{n-r}} & \mathbf{0} \\ \mathcal{K}_{\mathbf{0}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^{\mathbf{r}} \\ \mathbf{0} & \mathcal{L}_{\mathbf{0}} \end{bmatrix} \tag{2.43}$$

which simplifies to:

$$\begin{bmatrix} \mathbf{C}^{\mathbf{J}} \\ \mathbf{C}^{\mathbf{s}} \end{bmatrix} \left[\mathcal{K} - \boldsymbol{\varepsilon}_{\mathbf{s}}\mathcal{L}\right] = \begin{bmatrix} \mathcal{K} & \mathbf{0} \\ \mathbf{0} & \mathcal{L} \end{bmatrix} \tag{2.44}$$

Using the above definitions, the control-matrix equation $\mathbf{C^iE} = \mathbf{I}_n$ for the pathway shown in Fig. 2.1 can be expressed as[1]:

$$\begin{bmatrix} C_3^{J_3} & C_6^{J_3} & C_1^{J_3} & C_2^{J_3} & C_4^{J_3} & C_5^{J_3} \\ C_3^{J_6} & C_6^{J_6} & C_1^{J_6} & C_2^{J_6} & C_4^{J_6} & C_5^{J_6} \\ C_3^{s_1} & C_6^{s_1} & C_1^{s_1} & C_2^{s_1} & C_4^{s_1} & C_5^{s_1} \\ C_3^{s_2} & C_6^{s_2} & C_1^{s_2} & C_2^{s_2} & C_4^{s_2} & C_5^{s_2} \\ C_3^{s_4} & C_6^{s_4} & C_1^{s_4} & C_2^{s_4} & C_4^{s_4} & C_5^{s_4} \\ C_3^{s_5} & C_6^{s_5} & C_1^{s_5} & C_2^{s_5} & C_4^{s_5} & C_5^{s_5} \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & -\varepsilon_{s_2}^{v_3} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \left(\varepsilon_{s_6}^{v_6}\frac{s_5}{s_6} - \varepsilon_{s_5}^{v_6}\right) \\ \frac{J_3}{J_1} & \frac{J_6}{J_1} & -\varepsilon_{s_1}^{v_1} & 0 & 0 & 0 \\ \frac{J_3}{J_2} & \frac{J_6}{J_2} & -\varepsilon_{s_1}^{v_2} & -\varepsilon_{s_2}^{v_2} & 0 & 0 \\ 0 & 1 & 0 & -\varepsilon_{s_2}^{v_4} & -\varepsilon_{s_4}^{v_4} & 0 \\ 0 & 1 & 0 & 0 & -\varepsilon_{s_4}^{v_5} & \left(\varepsilon_{s_6}^{v_5}\frac{s_5}{s_6} - \varepsilon_{s_5}^{v_5}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.45}$$

2.4.2 Control pattern analysis

As previously discussed, control coefficients describe how sensitive metabolic variables are towards perturbations in reaction activities. Symbolic control coefficient expressions, such as those shown in Equations 2.30–2.32, take this a step further by additionally describing why control coefficients have certain values (as a result of their contributing components). In the aforementioned equations, for example, it is clear that if the value of $\varepsilon_{s_1}^{v_1}$ were 0 (due to reaction 1 being insensitive to its product S_1), then C_1^J would have a value of 1, and both C_2^J and C_3^J would be 0. Thus, these expressions allow for directly relating systemic properties to local properties.

From the demonstration in the previous section it should be clear that the control-matrix equation very effectively solves the problem of determining the control coefficients of a system from the elasticity coefficients, without the need for explicitly using the summation and connectivity theorems. This development has allowed metabolic modelling software to move away from the less accurate perturbation-based methods for calculating control coefficients that were prevalent at the time (e.g. [114, 115]), and that typically used perturbations of 1% or more, to matrix inversions that are accurate to within machine epsilon [116]. Indeed, some form of matrix inversion has been used in software packages since the early 1990s in order to calculate control coefficients (e.g. `Gepasi` [117]). The vast majority of implementations of the control-matrix equation, however, involve the numeric inversion of **E**. Thus symbolic

[1] A complete procedure for constructing this equation is given in Appendix A.

expressions of control coefficients in terms of elasticity coefficients are virtually absent from the literature for real metabolic systems.

A number of authors have, however, provided solutions to this problem. Hofmeyr [17, 118, 119] was one of the first to recognise the need for providing symbolic expressions for control coefficients. He therefore developed a diagrammatic method, called control pattern analysis, in which schemas of metabolic pathways (such as our example in Fig. 2.1) are used to generate symbolic control coefficient expressions by hand. Other authors, such as Thomas and Fell [22], and Schulz [120], have tackled this problem in a more conventional fashion and have provided software based solutions. All of these solutions, however, have some limitations: control pattern analysis, while less cumbersome and error prone than inverting matrices by hand, still requires considerable effort to carry out [17, 118, 119]; `MetaCon` [22] requires the user to manually choose a reference flux and produces different control coefficient expressions for different reference fluxes; and the software by Schulz [120] does not handle input files and can only be applied to linear pathways with a finite number of branches leading from the main pathway. Most recently, a software package called `SymCA` [23] was developed in our group in order to overcome the limitations of its predecessors and includes many concepts from control pattern analysis.

While control pattern analysis, as originally presented [17, 118, 119], is not ideally suited to generating control coefficient expressions, it does provide a useful framework for interpreting these expressions. Here, each of the terms of a control coefficient numerator corresponds to a so-called control pattern. Each control pattern, in turn, consists of a set of linked metabolite-enzyme pairs, which are represented by the elasticity coefficient factors of the numerator term. These control patterns, therefore, represent the physical routes through which a perturbation in an enzyme activity is transmitted towards the other enzymes in pathway, via changes in the linking metabolites [118]. To illustrate this principle we will once more turn to the 3-step pathway shown in Equation 2.23. Here a perturbation in the activity of reaction 1 is represented by C_1^J (Equation 2.31). Thus, reading the numerator term $\varepsilon_{s_1}^{v_2}\varepsilon_{s_2}^{v_3}$ from left to right, starting at s_1, we can interpret it to mean that the perturbation will affect the concentration of S_1, which in turn affects the activity of reaction 2, affecting S_2, and finally reaction 3. Using the shorthand qualitative notation suggested by Hofmeyr [17], and assuming an increase in v_1 activity, we can express this chain of effects as $\uparrow v_1 \uparrow s_1 \uparrow v_2 \uparrow s_2 \uparrow v_3$. In this example there is only a single control pattern, thus it represents the real order in which the effect of the enzyme modulation is propagated through the system [17]. In the case where there are multiple control patterns, or where there are multiple paths through a single pattern (as will be demonstrated below), the control pattern does not necessarily describe the real order in which an modulation effect will be propagated, however, it is still a useful construct

for understanding the overall effects that contribute towards the sensitivity of a metabolic variable towards a specific reaction [17].

As mentioned above, certain control coefficient numerators have multiple terms, and thus multiple routes through which a perturbation can propagate. Each of these routes can potentially carry a different weight in determining the ultimate control coefficient value. Here we will illustrate this principle using the pathway in Fig 2.1, focussing on the control patterns of $C_1^{J_1}$. The expression for this control coefficient is obtained through the inversion of the **E**-matrix shown in Equation 2.45:

$$C_1^{J_1} = \left(\frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_4} \varepsilon_{s_5}^{v_5}}{s_5} + \frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_4} \varepsilon_{s_6}^{v_6}}{s_6} + \frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_5} \varepsilon_{s_5}^{v_6}}{s_5} + \frac{J_6 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_4} \varepsilon_{s_4}^{v_5} \varepsilon_{s_5}^{v_6}}{s_5} - \frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_4} \varepsilon_{s_6}^{v_5}}{s_6} - \frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_4} \varepsilon_{s_5}^{v_6}}{s_5} - \frac{J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_5} \varepsilon_{s_6}^{v_6}}{s_6} - \frac{J_6 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_4} \varepsilon_{s_4}^{v_5} \varepsilon_{s_6}^{v_6}}{s_6} \right) / \Sigma \tag{2.46}$$

where each term in the numerator represents a separate control pattern[2]. Numbering the control patterns from left to right, control pattern 1 (CP1) is $J_3 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_3} \varepsilon_{s_4}^{v_4} \varepsilon_{s_5}^{v_5} / s_5$ and CP8 is $-J_6 \varepsilon_{s_1}^{v_2} \varepsilon_{s_2}^{v_4} \varepsilon_{s_4}^{v_5} \varepsilon_{s_6}^{v_6} / s_6$. In the case of CP2 and CP6 we can see that these two control patterns differ only by two factors: CP2 contains $\varepsilon_{s_6}^{v_6} / s_6$, whereas CP6 contains $-\varepsilon_{s_5}^{v_6} / s_5$. Thus, any observed difference in the values of these two patterns can be traced to the values of these factors.

It is worth noting that patterns can also be scaled by fluxes and metabolite concentrations. These scaling factors depend on the path represented by a control pattern. Whenever a control pattern passes through a metabolic branch other than the branch containing the flux of interest (or reference flux), and includes all the metabolite-enzyme pairs of that branch, it will be scaled by the flux of that branch [17, 23, 118]. This can be seen in the flux scaling factor J_3 contained in both CP2 and CP6. In the case where a control pattern includes a path passing through a metabolite-enzyme pair where the metabolite is part of a moiety conserved cycle, the control pattern will be scaled by the concentration of that metabolite [23]. In CP2 the metabolite scaling factor is s_6, while in CP6 it is s_5. These scaling factors are not critical for tracing the chains of local effects, but do contribute to the overall value of a control pattern and are a result of the flux relationships between different branches and of expressing control patterns in the "standard" form originally defined by Hofmeyr [121] rather than using nested block elasticities.

[2]For the purpose of this illustration the denominator expression is unnecessary, therefore it is represented by Σ for the sake of clarity. The full denominator term consists of all of the control patterns of all of the flux control coefficients of the system.

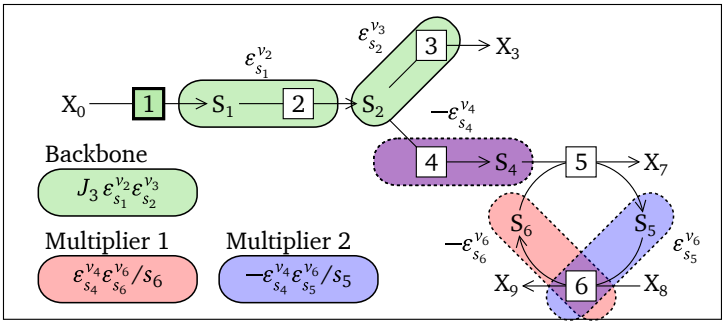


Figure 2.2: Control patterns of the pathway shown in Fig. 2.1. Here the backbone pattern belonging to CP2 and CP6 is highlighted in green. The red multiplier pattern (1) belongs to CP2, whereas the blue multiplier pattern (2) belongs to CP6. The purple metabolite-enzyme pair represented by $\varepsilon_{s_4}^{v_4}$ is common to both CP2 and CP6.

In order to clarify the control patterns further, they can subdivided into smaller backbone and multiplier patterns [118]. In the case of flux control coefficients, backbone patterns are defined as the uninterrupted chains of local effects that originate from the modulated enzyme and connect two terminal metabolites. In the case of metabolic cycles, the backbone is also permitted to loop back onto itself [118]. Additionally, the backbone pattern passes through the reference flux. For concentration control coefficients, the backbone represents the path that links the modulated enzyme and the controlled metabolite. A single control coefficient can have multiple backbones. Multiplier patterns, on the other hand, are not directly linked to the backbone pattern and only occur in the case where there are branches in a pathway. Thus, any single backbone pattern can be combined with various multiplier patters to produce different control patterns. Moreover, the same multiplier patterns can be found in control patterns with different backbones. In Fig 2.2, CP2 and CP6 are shown on the same diagram. Here it is clear that these control patterns share the same backbone $J_3\varepsilon_{s_1}^{v_2}\varepsilon_{s_2}^{v_3}$, while CP2 has the multiplier pattern $\varepsilon_{s_4}^{v_4}\varepsilon_{s_6}^{v_6}/s_6$ and CP6 the pattern $-\varepsilon_{s_4}^{v_4}\varepsilon_{s_5}^{v_6}/s_5$. By breaking the chains of local effects into smaller parts it becomes much easier to parse and analyse the control structure of the pathway; however, as with the scaling factors, multiplier patterns emerge as a result of expressing control patterns in the standard form, and thus do not represent the real order in which effects are propagated in a system [17].

In general, symbolic expressions of control coefficients have a variety of advantages over their numeric counterparts. As shown above they allow for the systematic dissection of the control structure of a metabolic pathway [17, 23, 118]. They make it possible to quantify the effects of different components of the full control structure under various different conditions, and to draw conclusions regarding the influence that these components have on the control properties [23]. Due to the fact that the control coefficient expressions depend only

on the knowledge of the pathway stoichiometry and the presence of allosteric regulatory interactions, some general conclusions regarding the control structure can be drawn even in the absence of a full kinetic characterisation [23]. Ultimately, symbolic control coefficients and control patterns relate the control in the system to the properties and interactions of its components in a way that is impossible with numeric control coefficients [17, 23, 118].

2.4.3 Symbolic control analysis in metabolic studies

While a variety of solutions exists for performing symbolic control analysis (at least to some degree), there are very few examples of the application of this technique in the manner described above. For example, in two studies by the original authors of `MetaCon` where symbolic control coefficient expressions were generated [27, 122], these expressions were predominantly used to calculate numeric control coefficient values from elasticity coefficients. These authors substituted elasticity coefficient values obtained under a number of different conditions into the control coefficient expressions, thus determining the control coefficient values for these different conditions. However, instead of obtaining new values for each of the pertinent elasticity coefficients under the different conditions, these studies only considered some elasticity coefficients and assumed that the others remained constant regardless of the external conditions, thus resulting in arguably unrealistic control coefficient values (as noted by the authors themselves).

The first example of the application of symbolic control analysis in the manner described in Section 2.4.2, is in the thesis by Akhurst [24] on the development of the original version of `SymCa`. Here, symbolic control analysis and control-pattern analysis were performed on a variety of models of real metabolic systems. While none of these case studies have been formally published, in our opinion, they are important examples of the potential of this framework. In one of these studies, for instance, Akhurst investigated the control patterns responsible for the shift in control under different pH and environmental conditions in models of fermentation in *Saccharomyces cerevisiae* [87]. He demonstrated that an increase in the glucose uptake rate observed in immobilised cells was a result of the "turning on" of a part of the system related to carbohydrate production [24]. In a case study of a model of sucrose accumulation in sugar cane (*Saccharum officinarum*) culm [123], he showed that decreased futile cycling in maturing internodes is related to a decrease in control of control patterns acting via neutral invertase [24].

2.5 Metabolic supply-demand analysis

Metabolic supply-demand analysis is a framework for investigating the regulation, behaviour and control of metabolic pathways [14]. This framework stemmed from work carried out by Hofmeyr and Cornish-Bowden during the 1990s [25, 37, 107, 110] and was developed in response to these authors' perception of a widespread disregard for the greater cellular context of metabolic pathways during metabolic studies. They argued that the concepts of function, regulation, and organisation of the metabolic network (as introduced in Section 2.1) are essential for understanding the behaviour of metabolism. Here metabolism is, therefore, considered to be a molecular economy that is functionally organised into blocks of reactions linked by a common intermediate or a moiety-conserved cycle. In keeping with this economic metaphor, a block tasked with producing the linking intermediate is called the supply block, whereas the consuming block is called the demand block. According to these authors, an example of a well regulated system is one in which the flux can sensitively respond to external stimuli, such as an increase in demand for a product, while maintaining constant intermediate concentrations [37]. Therefore, one of the main goals of SDA is to illustrate and quantify the differentiation of the supply and demand blocks in terms of these functions.

One shortcoming of supply-demand analysis is that it requires a researcher to choose the intermediate which divides the metabolic system into supply and demand blocks [14, 15]. This is problematic, because the functional organisation of a metabolic system is frequently unknown, unclear, or ambiguous, and in some cases metabolic function as a whole is very difficult to identify [3]. Any selection of a linking metabolite may therefore introduce a bias. To overcome this limitation Rohwer and Hofmeyr [15], have generalised supply-demand analysis so that each metabolite is, in turn, automatically selected as the linking metabolite using an automated computational method. In this variant of supply-demand analysis there are as many supply and demand blocks as there are intermediates and moiety-conserved cycles. This method overcomes the selection bias by including the identification of functional differentiation as part of the analysis, rather than presupposing metabolic function. Moreover, GSDA can be used to identify and quantify additional regulatory features that cannot be detected using the single linking metabolite treatment of SDA, as will be discussed below in Section 2.5.2. Both Sections 2.5.1 and 2.5.2 are largely summaries of the ideas presented in the original publications [14, 15].

2.5.1 Quantitative analysis of supply-demand systems

This section will commence with a description of the procedure of ordinary supply-demand analysis using the linear metabolic pathway depicted in Fig 2.3 as an example. In this 5-step

pathway the intermediate S_2 acts is the link between the supply block, consisting of reactions 1 and 2, and the demand block, consisting of reactions 3 to 5. Note that the supply and demand blocks are completely isolated from one another, save for their interaction via S_2, which is considered as parameter (see Section 2.3.1) to both reaction blocks.

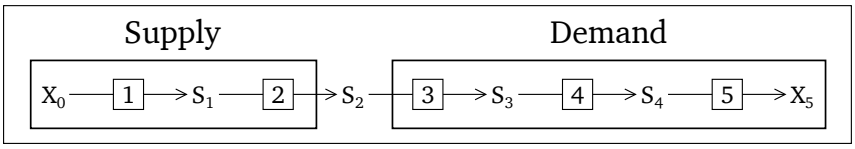


Figure 2.3: A simple 5-step linear metabolic pathway. Here S_2 is considered the linking metabolite. Therefore the supply block consists of reactions 1 and 2, whereas the demand block consists of reactions 3–5.

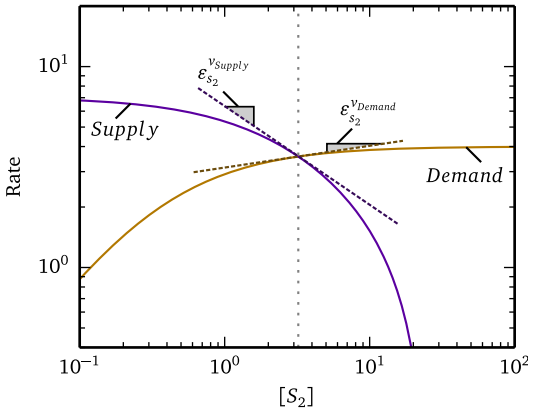


Figure 2.4: A rate characteristic plot for the system shown in Fig. 2.3. The supply and demand curves for S_2 are plotted in double logarithmic space. Block elasticity coefficients are shown as tangents to the rate curves (dashed lines). This figure was adapted from [14].

Rate characteristic plots [25] are one of the main tools of SDA. These plots are visual representations of the steady-state properties of the supply and demand blocks and show how flux and concentration control are distributed among these blocks. Here the natural logarithms of the supply and demand block rates are plotted against the natural logarithm of the linking metabolite concentration as demonstrated in Fig. 2.4. The depicted rate curves show the local fluxes of the isolated reaction blocks in response to changing linking intermediate concentration. The point where these curves intersect indicates the steady-state flux and linking intermediate concentration for the whole, undivided, system. Finally, the elasticity coefficients of the reaction blocks towards the linking intermediate are indicated

as slopes to the rate curves at their point of intersection. These graphs can be generated *in silico* by parametrising the linking metabolite and varying its concentration over a range, or through application of the double-modulation method [124] on an experimental system [28, 29, 125, 126].

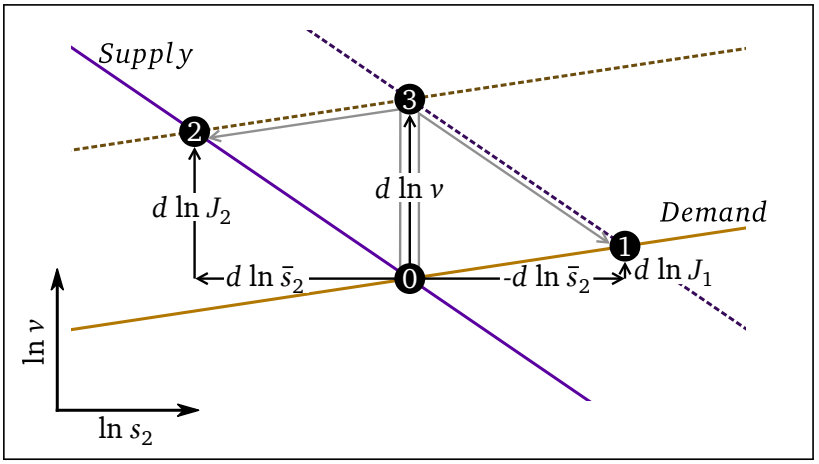


Figure 2.5: A close-up view of the steady state shown in the rate characteristic of Fig. 2.4. The reference steady state is indicated by ⓪. Here a small perturbation of $d \ln v$ in the supply or demand activities lead to new steady states respectively indicated by ① and ②. Perturbing both blocks by $d \ln v$ results in the the steady state at ③. This figure was adapted from [14].

The second important tool of SDA is metabolic control analysis [8, 9]. Using the elasticity coefficients shown in Fig. 2.4, we can demonstrate the effect of a perturbation of either the supply or the demand block activities on the steady state. In Fig. 2.5, point 0 indicates the reference steady state shown in Fig. 2.4, whereas points 1 and 2 indicate two new steady states that are reached after respectively increasing supply and demand activities by the same amount $d \ln v$. At point 2 the flux and the steady-state concentration of S_2 have respectively increased by $d \ln J_2$ and decreased by $d \ln \bar{s}_2$ compared to the reference state, while at point 1 they have increased respectively by $d \ln J_1$ and $d \ln \bar{s}_2$. The steady state indicated by point 3 is the result of an increase in the activities of both the supply and demand blocks by $d \ln v$; here the flux has increased by $d \ln J_1 + d \ln J_2 = d \ln v$, while there is no change in S_2 concentration from the reference state. The ratios of the change in steady-state flux over the activity change, and change in steady-state concentration over the activity change, are respectively the flux control coefficients:

$$C^J_{supply} = \frac{d \ln J_1}{d \ln v_{supply}}, \quad C^J_{demand} = \frac{d \ln J_2}{d \ln v_{demand}} \tag{2.47}$$

and the concentration control coefficients:

$$C^{s_2}_{supply} = \frac{d\ln s_2}{d\ln v_{supply}}, \quad C^{s_2}_{demand} = \frac{-d\ln s_2}{d\ln v_{demand}} \tag{2.48}$$

Using these control coefficient expressions together with the elasticity coefficients shown in Fig. 2.3, the previously discussed summation (Equations 2.17 and 2.19) and connectivity theorems (Equations 2.21 and 2.22) can be derived. These theorems can, in turn, be used to express the control coefficients in terms of the elasticity coefficients of the supply and demand blocks. The flux control coefficients are therefore:

$$C^{J}_{supply} = \frac{\varepsilon^{v_{demand}}_{s_2}}{\varepsilon^{v_{demand}}_{s_2} - \varepsilon^{v_{supply}}_{s_2}} \tag{2.49}$$

and

$$C^{J}_{demand} = \frac{-\varepsilon^{v_{supply}}_{s_2}}{\varepsilon^{v_{demand}}_{s_2} - \varepsilon^{v_{supply}}_{s_2}}, \tag{2.50}$$

whereas the concentration control coefficients are:

$$C^{s_2}_{supply} = -C^{s_2}_{demand} = \frac{1}{\varepsilon^{v_{demand}}_{s_2} - \varepsilon^{v_{supply}}_{s_2}} \tag{2.51}$$

This leads us to one of the most powerful applications of supply-demand analysis: the ability to detect and illustrate the functional differentiation of reaction blocks in metabolic systems. From Equations 2.49 and 2.50 we can see that the ratio $|\varepsilon^{v_{supply}}_{s_2}/\varepsilon^{v_{demand}}_{s_2}|$ determines the distribution of flux control between the two blocks. In the case where this ratio is larger than 1 the demand block has more control over flux, while if it is less than 1 the supply block has more control over flux. According to the summation theorem, it is clear that as one block gains more control, the other loses control. In the case of the concentration control coefficients, both blocks have the same amount of control regardless of the values of the elasticity coefficients. However, as the value of the denominator of Equation 2.51 increases there is a decrease in the concentration control of S_2 by both blocks and therefore an increase in homoeostasis of this metabolite. Here it is clear that the reaction block with the numerically larger elasticity will contribute more towards increased homoeostasis than a block with a small elasticity.

Taking these relationships between the control and elasticity coefficients together we can conclude that as one reaction block gains control over flux it loses influence over the degree of homoeostasis of the linking metabolite, while the other block loses flux control and increases its influence over homoeostasis. Flux control and homoeostasis are therefore mutually exclusive functions. The two extremes of this principle are illustrated in Fig. 2.6. Here

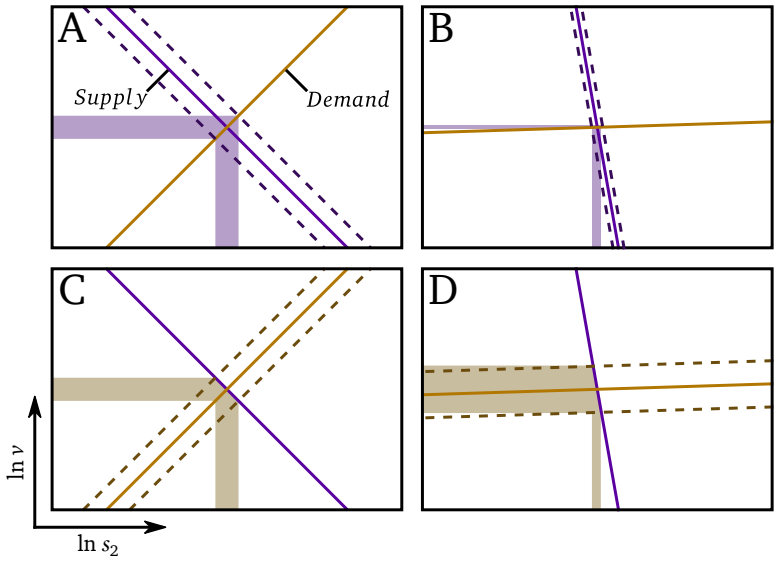


Figure 2.6: The effect of different block sensitivities on functional differentiation. The left-hand side (**A** and **C**) shows a system where both supply and demand blocks are equally sensitive to a perturbation in the linking metabolite, whereas the right-hand side (**B** and **D**) shows a system where the demand block is almost completely insensitive to the linking metabolite, while the supply block is extremely sensitive. Supply activity perturbations are indicated by **A** and **B**, while demand activity perturbations are indicated by **C** and **D**. Shaded areas indicate the magnitude of the change in steady-state variables resulting from a perturbation. All the indicated perturbations are equal in magnitude. This figure was adapted from [14].

A and C show the case where both reaction blocks have elasticity coefficients that are equal in magnitude. Perturbing either of the supply (A) or the demand activities (C) in this system by the same amount results in new steady states where the magnitudes of the changes in the steady-state variables are equal; both blocks have equal flux control and equally determine the degree of homoeostasis. On the other hand, B and D show the case where $\varepsilon_{s_2}^{v_{supply}}$ has an extremely large negative value and $\varepsilon_{s_2}^{v_{demand}}$ is close to zero. Here a perturbation in supply activity (B) has almost no effect on the steady-state flux, while a perturbation in demand activity (D) has a large effect on the steady-state flux and leads to a significantly higher or lower flux value. Perturbing either of these reaction blocks has little effect on the steady-state S_2 concentration. In this example demand has virtually complete control over flux, while the degree of homoeostasis is largely determined by supply.

2.5.2 Generalising supply-Demand analysis

As previously mentioned, GSDA generates rate characteristic plots for each intermediate in a system. This removes the restriction that reaction blocks should only be able to communi-

cate through the linking intermediate. These changes enable new avenues for investigation; however, they also call for certain adjustments to how systems are treated.

To illustrate these adjustments we will once more use the 5-step system shown in Fig. 2.3. In Fig. 2.7 a rate characteristic plot for the supply and demand blocks of S_2 is shown. While mostly similar to Fig. 2.4, here the block elasticity coefficients have been replaced with equivalent response coefficients. Additionally, the elasticity coefficients of the supply and demand reactions towards the linking metabolite are included and are also represented as lines intersecting the steady-state point. These elasticity coefficients are important for identifying sites of regulation and regulatory metabolites, as will be discussed in Section 2.5.3.2. Another change stems from the allowance for multiple routes of interaction between reaction blocks. In the case where such additional routes exist between two blocks, the control coefficient expressions of Eqs. 2.49-2.51 are invalidated. In these cases rate characteristic plots may also include lines representing partial response coefficients, as demonstrated in Section 2.5.3.3.

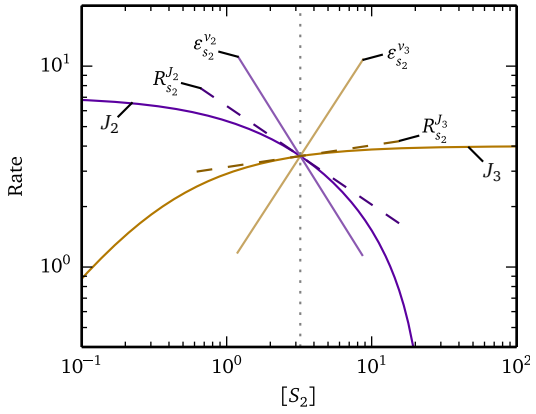


Figure 2.7: A generalised supply-demand analysis compatible rate characteristic plot for the system shown in Fig. 2.3. As in Fig. 2.4, the supply and demand curves for S_2 are plotted in double-logarithmic space. Here, block elasticity coefficients are replaced with equivalent block response coefficients, which are indicated as dashed tangents to the rate curves. Elasticity coefficients of the linking reactions towards S_2 are shown as solid lines which intersect the point representing the steady state. This figure was adapted from [15].

2.5.3 Interpretation of rate characteristics in generalised-supply demand analysis

The rate characteristic plots generated in GSDA can be interpreted on multiple levels, thereby revealing several regulatory features. In addition to illustrating and quantifying functional

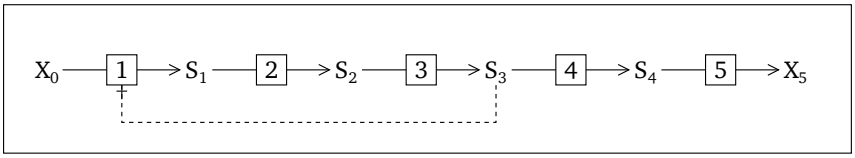


Figure 2.8: A simple 5-step linear metabolic pathway with a feedback loop. Here S_3 inhibits reaction 1. In generalised supply demand analysis, each metabolite will, in turn, be treated as a linking metabolite, resulting in as many supply and demand block combinations.

differentiation as in SDA, GSDA can also be used to identify potential sites of regulation, regulatory metabolites and the relative regulatory importance of different routes of interaction between the linking intermediate (or moiety-conserved cycle) and a reaction block.

In this section we will explore these features using the pathway shown in Fig. 2.8. This 5-step pathway is essentially the same as the one in Fig. 2.3; however, here a negative feedback loop from S_3 to reaction 1 has been introduced. We will consider the rate characteristic plots of the reaction blocks of S_2, S_3 and S_4.

2.5.3.1 Differences in rate characteristic shapes

The first level of inspection simply involves the shapes of the different rate characteristics of a system without considering the response and elasticity coefficients. In a series of rate characteristic plots depicting the supply and demand curves for different intermediates of the same system, a change in the shapes of the rate characteristics when moving from one intermediate to the next potentially indicates a site of regulation due to a change in kinetics.

An example of such a case is shown in Fig. 2.9. Here the rate characteristics for the supply and demand blocks for S_3 (Fig. 2.9A) and for S_4 (Fig. 2.9B) are shown and can be seen to have notably different shapes. On the other hand the rate characteristic plots for S_1 and S_2 (which are not shown here) have supply and demand curves that are nearly identical to those of S_3. It is therefore clear that the difference between the rate characteristic shapes of S_3 and S_4 is due to the kinetic properties of the enzyme of reaction 4. In this case the enzyme catalysing reaction 4 is insensitive towards its product S_4 at the steady state.

2.5.3.2 Comparisons of elasticity and response coefficients

The second level of inspection involves the comparison of elasticity coefficient of the supply and demand enzymes with their corresponding block response coefficients for each metabolite in the system. Here an agreement between these two coefficients can point to sites of regulation, regulatory metabolites, or points of functional differentiation.

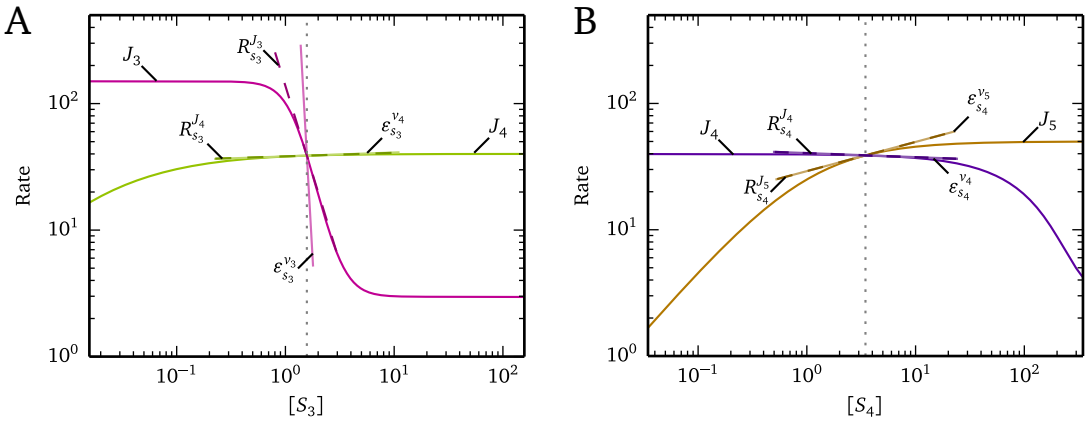


Figure 2.9: Rate characteristic plots of S_3 and S_4 for the pathway in Fig. 2.8. **A** indicates the supply and demand curves of S_3, while **B** indicates the supply and demand curves of S_4. The difference in rate characteristic shapes between **A** and **B** is due to the in insensitivity of reaction 4 towards S_4. Dashed tangents indicate response coefficients, whereas solid lines passing through the steady-state point indicate elasticity coefficients. This figure was adapted from [15].

According to the partitioned response property (Equation 2.14) when $R^J_{s_j} = \varepsilon^{v_i}_{s_j}$ then $C^J_i = 1$. This signifies that the enzyme that interacts directly with the linking intermediate has complete control over the flux of its reaction block. It also means that any effect that the linking metabolite has on the enzyme will be directly transmitted to the rest of the reaction block; the linking metabolite can, therefore, be classified as a "regulatory metabolite" with respect to the reaction block in which this condition is satisfied. In Fig. 2.9B, $R^{J_4}_{s_4}$ equals $\varepsilon^{v_4}_{s_4}$, therefore reaction 4 has full control over the supply block flux and S_4 is a regulatory metabolite.

In the case where a supply or demand enzyme is completely insensitive towards the linking metabolite, it also results in an agreement between the elasticity and response coefficients. Here $R^J_{s_j} = \varepsilon^{v_i}_{s_j} \approx 0$, therefore the value of the control coefficient is irrelevant. This scenario indicates that the system is functionally differentiated at this point, as discussed in Section 2.5.1.

Notably, there will always be an agreement between the elasticity and response coefficients if a reaction block consists of a single enzyme. This represents the trivial case and will always be encountered for the first and last intermediates of a pathway. The agreement between $R^{J_5}_{s_4}$ and $\varepsilon^{v_5}_{s_4}$ in Fig. 2.9B is an example of this scenario.

2.5.3.3 Multiple routes of regulation

Finally, in the case where there are multiple routes of interaction between a linking metabolite and a reaction block, GSDA can be used to quantify the relative regulatory importance of these routes. Here we make use of visual representations of partial response coefficients. For the supply block of S_3 in Fig. 2.8 the total response coefficient can be expressed in terms of its partial response coefficients (as per Equations 2.12 and 2.13) as follows:

$$R^{J_3}_{s_3} = {}^{v_1}R^{J_3}_{s_3} + {}^{v_3}R^{J_3}_{s_3} \tag{2.52}$$

Each of the partial response coefficients quantifies the contribution towards the total response of the interaction that it represents.

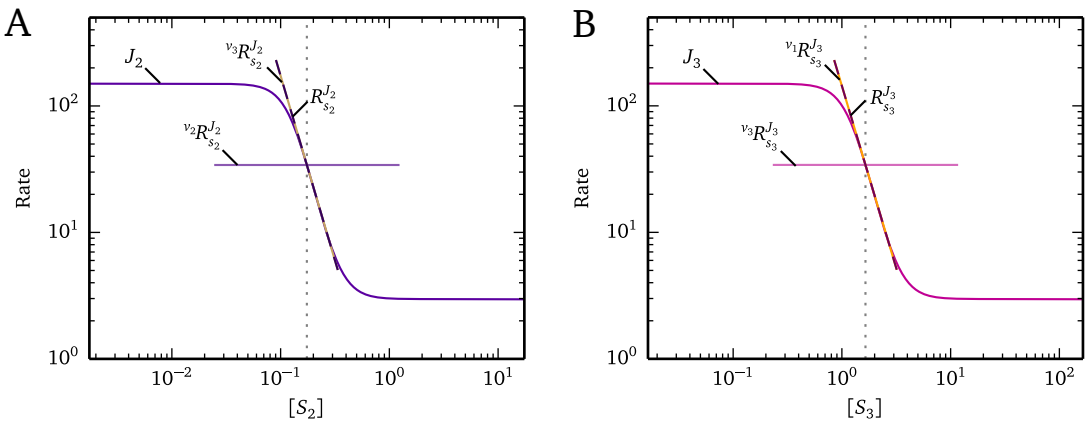


Figure 2.10: Rate characteristic plots of S_2 and S_3 for the pathway in Fig. 2.8. Here only the respective supply curves of S_2 (**A**) and S_3 (**B**) are shown. Partial response coefficients, shown as solid lines passing through the steady-state point, indicates the relative contributions of different routes of regulation towards the total response coefficients shown as dashed lines. This figure was adapted from [15].

An example of this analysis is given in Fig. 2.10 where the supply block fluxes for S_2 (Fig. 2.10A) and S_3 (Fig. 2.10B) are shown together with lines representing total and partial response coefficients. In the case of the S_3 supply block, the two partial response coefficients represent the product inhibition of reaction 3 ($^{v_3}R^{J_3}_{s_3}$) and the allosteric inhibition of reaction 1 ($^{v_1}R^{J_3}_{s_3}$) by S_3. Here feedback inhibition is clearly solely responsible for the supply flux response, while product inhibition has no regulatory effect. The metabolite S_2 can similarly interact with its supply block via two different routes. The first route is simply through product inhibition of reaction 2 ($^{v_2}R^{J_2}_{s_2}$). The second route, however, is more complicated and takes place through the demand block ($^{v_3}R^{J_2}_{s_2}$); S_2 interacts with reaction 3, thus affecting the

concentration of S_3 which, in turn, affects the activity of reaction 1. Fig. 2.10A shows that this is the dominant regulatory route.

2.5.4 Supply-demand analysis in metabolic studies

Since its inception, supply-demand analysis has been used in a variety of experimental studies, many of which have confirmed the idea that flux control resides in the demand for certain intermediates. The earliest example of a study in which the techniques of SDA were applied is that by Koebmann *et al.* [125], where they performed an experimental SDA by modulating the ATPase activity in *Echerichia coli*. In this organism the additional ATPase activity resulted in a 70% increase in the glycolytic flux, leading to the conclusion that this pathway is predominantly controlled by ATP demand. In another example, Jørgensen *et al.* [127] modulated the supply of CTP in *Lactococcus lactis* by altering the expression level of the gene (*pyrG*) encoding for its producing enzyme CTP synthase. They found that decreasing the level of CTP synthase to 43% had no effect on the growth rate, indicating that flux control does not lie in the supply-block. Thus, while this enzyme is responsible for determining the degree of homoeostasis of CTP concentration, CTP synthase was shown to have a high degree of control on CTP concentration, therefore indicating a low degree of CTP homoeostatic maintenance. As demonstrated by the relatively recent study of Arsac *et al.* [128], this phenomenon is not limited to bacterial systems. Through an *in vivo* modular control analysis of a mouse muscle model, these authors showed that ATP demand in muscle fibres is responsible for muscle contraction. Additionally, they concluded that the persistent side effects of bipuvicaine injections (which are used as a neurotransmission blocker) on muscle energetics is due to it decreasing the demand for ATP in muscle fibres.

In addition to studies that have made explicit use of the techniques of this framework, the idea of a demand-controlled flux has also been used to make sense of previously published results. McCormick *et al.* [26] have, for instance, used this framework to argue that photosynthetic rates in sugercane leaves are determined by the demand for carbon from sink tissues. Similarly, in an extensive review, Morandini [129] used numerous published sources to argue that the flux control for several "building blocks" of metabolism, such as amino acids and lipids, are controlled by their demand. For example, he attributed previous failed attempts to increase lipid flux to the modulation of supply enzymes thought to be rate limiting by the authors of these studies (e.g. [130, 131]), while giving examples of successful attempts to increase lipids in which the demand was modulated (e.g. [132, 133]).

There have, however, been contrary results that indicate a predominantly supply-controlled flux. In another study, Koebmann *et al.* [29] compared glycolytic flux control by ATPase in *Echerichia coli* and *Lactococcus lactis*. In contrast to *E. coli*, they found that the flux of glycoly-

sis is supply-controlled in *L. lactis*; however, they concluded that this was due to the glycolytic flux already operating near its maximal capacity in this organism prior to the increase in ATPase (thus mirroring a scenario that would occur when the J_4 curve in Fig. 2.9A is increased so that it intersects the J_3 curve at $S_3 \lesssim 6 \times 10^{-1}$). Similarly, Schafer *et al.* [134] have shown that the control of flux towards glycogen in muscle can be attributed to the GLUT4 glucose transporter in the supply block for glucose-6-phosphate.

While SDA has clearly yielded many important results in numerous studies, the predominantly computational technique of generalised supply-demand analysis has, to our knowledge, only been utilised in a single published study [31]. This study on tracing the regulatory routes of metabolic pathways is presented as Chapter 4 of this dissertation.

2.6 Distinguishing between thermodynamic and kinetic effects in enzyme-catalysed reactions

As already discussed in Section 2.1, enzymes are central components in the regulation of metabolism. In addition to their crucial function of increasing the rate of reactions far beyond what would otherwise be possible in the mild conditions of the cell, they also counteract or augment the existing mass-action trend in a way the satisfies the cell's requirements [25]. Additionally, enzymes act as control knobs that can be tweaked in order to regulate the flow of matter through the cell [25]. In support of these goals enzymes have evolved a wide array of features [25]: they have high binding and catalytic specificity, there are a variety of mechanisms for binding substrates and products, they can exhibit cooperative binding, and they have the ability to be affected by allosteric signals.

The degree to which reactions are displaced from equilibrium has long been thought to play a role in determining the behaviour of metabolic pathways. Here the value of the disequilibrium ratio Γ/K_{eq} (or ρ) has been used as measure of this distance. Classically, far-from-equilibrium reactions were believed to be so-called rate-limiting steps, which are thought to dominate the control of flux. Near-equilibrium reactions were thought to have little potential for controlling flux [12]. These views are, however, flawed in a number of ways. As we have discussed and clearly demonstrated in Sections 2.3 and 2.4, metabolic control is typically shared between the reactions in a metabolic pathway. Even in one of the original papers on metabolic control analysis, Kacser and Burns [8] demonstrated, by deriving control coefficients in terms of the disequilibrium ratios of the steps in a linear pathway, that distance from equilibrium is not an accurate measure for determining control. Another flaw is that there has historically been little theoretical justification of what exactly it means to be near-equilibrium or far-from-equilibrium [12, 16]. A commonly excepted definition is the

one put forth by Rolleston [135] in 1972, in which reactions with $\rho \leq 0.05$ are regarded as far-from-equilibrium and those with $\rho \geq 0.2$ as near-equilibrium. This suggestion has, however, been criticised as being inadequate in terms of its intended purpose [12, 16].

These criticisms, however, are not to say that distance from equilibrium is irrelevant, or that there is no basis for the aforementioned claims. Distance from equilibrium is indeed a valid measure for predicting how strongly a reaction rate can be affected by the regulatory properties of a enzyme (such as binding or allosteric effects) as near-equilibrium reactions will tend to be predominantly thermodynamically controlled, i.e., dominated by the mass-action term in the rate equation [25]. The balance between the mass-action driving force and the effect of the enzyme also has important implications for the homoeostatic maintenance of intermediate concentrations: if the reaction has a large negative value for $\Delta G^{\circ\prime}$ reactions have to be kept far away from equilibrium in order to prevent product accumulation, while reactions with $\Delta G^{\circ\prime}$ values close to 0 suffer from substrate accumulation if they are not allowed to approach equilibrium [136]. However, the distance that a reaction is from equilibrium does not reveal any information regarding the actual effect of intrinsic mass-action or enzyme binding on the reaction rate; it merely indicates the potential for being controlled by either of these components.

In order to understand the regulatory effects of enzymes, it is therefore imperative that we separate their contributions from the intrinsic mass-action properties of the catalysed reaction. By building on previous work [25], Rohwer and Hofmeyr [16] have developed a method in which rate equations are rewritten in such a way that this goal is achieved in two ways: in the first they distinguish between far-from-equilibrium and near-equilibrium reactions and quantify the degree of thermodynamic and kinetic control of reaction rate, and in the second they quantify the intrinsic mass-action contribution towards reaction rate as well as the degree to which it is counteracted or augmented by the enzyme [16]. Below follows a summary of this framework.

2.6.1 Thermodynamic contribution and distance from equilibrium

In the method developed by Rohwer and Hofmeyr [16], ρ is also used to discriminate between near-equilibrium and far-from-equilibrium reactions. Here, however, the boundaries which define these two classifications are based on the degree to which the kinetic and thermodynamic aspects of reactions determine the reaction rate at different substrate concentrations. Therefore, in the first step of this procedure the rate equation of any reversible reaction is expressed in a way that separates the kinetic and thermodynamic aspects. For example, the

bi-bi mass-action reaction:

$$mA + nB \rightleftharpoons pC + qD \tag{2.53}$$

its rate equation:

$$v = k_f a^m b^n - k_r c^p d^q \tag{2.54}$$

can be rewritten as:

$$\begin{aligned} v &= k_f a^m b^n \left(1 - \Gamma/K_{eq}\right) \\ &= k_f a^m b^n \times (1-\rho) \end{aligned} \tag{2.55}$$

where the concentrations of the species A, B, C, and D are raised to their respective stoichiometric coefficients m, n, p and q; k_f and k_r respectively represent the forward and reverse rate constants; and K_{eq} is obtained through the relationship $K_{eq} = k_f/k_r$. Here the forward rate term $v_f = k_f a^m b^n$ represents the kinetic contribution towards the rate, whereas the term $v_{(1-\rho)} = 1-\rho$ describes the thermodynamic contribution. In this form, however, v_f and $v_{(1-\rho)}$ are multiplicative and do not allow for the clear discrimination between their respective effects, thus we take the logarithm of Equation 2.55:

$$\ln v = \ln v_f + \ln v_{(1-\rho)} \tag{2.56}$$

which converts it to an expression with additive terms.

Now the expression in Equation 2.56 can be used to visualise the contributions of kinetic and thermodynamic functions towards the reaction rates for different substrate and product concentrations as shown in Fig. 2.11A and B, respectively. From Fig. 2.11A it is clear that for concentrations far from equilibrium (where equilibrium is encountered when $a = 5$) $\ln v_f$ is dominant, whereas near equilibrium the term $\ln v_{(1-\rho)}$ dominates. Additionally, there is a large range of values for which both terms simultaneously make a noteworthy contribution towards $\ln v$. In Fig. 2.11B the $\ln v_f$ contribution remains constant as this term does not account for changes in product concentration.

The degree of either kinetic or thermodynamic control can thus be quantified in terms of the fractional difference between the slope of the $\ln v$ and the slope of either of $\ln v_f$ or $\ln v_{(1-\rho)}$. These slopes are represented by the elasticity coefficients obtained from partial differentiation of the rate function with respect to $\ln a$:

$$\varepsilon_a^v = \frac{m}{1-\rho}, \quad \varepsilon_a^{v_f} = m, \quad \varepsilon_a^{v_{(1-\rho)}} = \frac{m\rho}{1-\rho} \tag{2.57}$$

and with respect to $\ln c$:

$$\varepsilon_c^v = \frac{-p\rho}{1-\rho}, \quad \varepsilon_c^{v_f} = 0, \quad \varepsilon_c^{v_{(1-\rho)}} = \frac{-p\rho}{1-\rho} \tag{2.58}$$

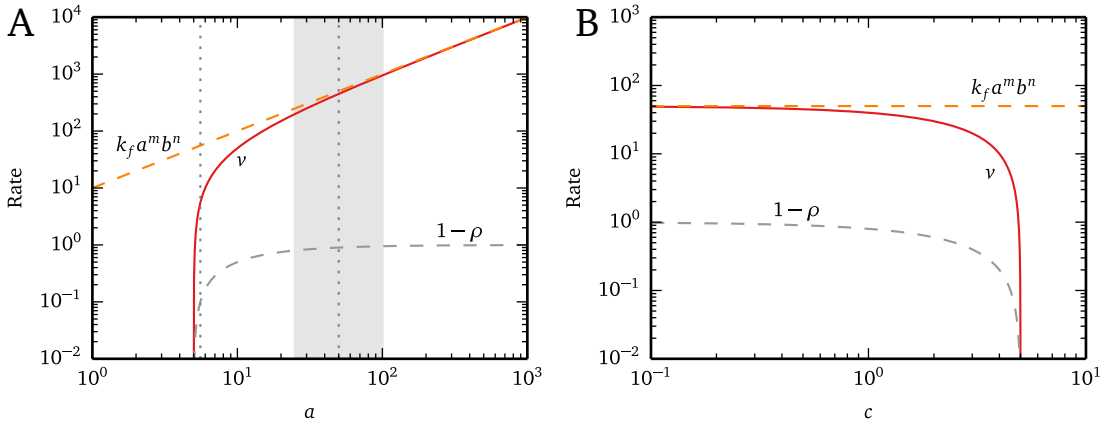


Figure 2.11: The rate of the bi-bi mass action reaction in Equation 2.53 as a function of substrate and product concentration. In A the terms $k_f a^m b^n$ and $1-\rho$, which constitute ν, are plotted together with ν in double-log space as a function of a, whereas B shows the same terms as a function of c in double-log space. In both A and B $k_f = 10$, whereas K_{eq}, b, d, m, n, p, and q are set to 1. In A $c = 5$, while in B $a = 5$. The vertical dotted lines in A correspond, from left to right, to the values of a where $\rho = 0.9$ and $\rho = 0.1$, whereas the shaded area indicates $0.05 < \rho < 0.2$. This figure was adapted from [16].

The fractional differences between the slopes of the logarithmic functions are thus given by the ratios of elasticity coefficients:

$$\frac{\varepsilon_a^{\nu} - \varepsilon_a^{\nu_f}}{\varepsilon_a^{\nu}} = \frac{\varepsilon_a^{\nu_{(1-\rho)}}}{\varepsilon_a^{\nu}} = \rho \tag{2.59}$$

and

$$\frac{\varepsilon_a^{\nu} - \varepsilon_a^{\nu_{(1-\rho)}}}{\varepsilon_a^{\nu}} = \frac{\varepsilon_a^{\nu_f}}{\varepsilon_a^{\nu}} = 1-\rho \tag{2.60}$$

where the Equation 2.59 is the discriminator for kinetic control and Equation 2.60 is the discriminator for thermodynamic control.

Now a reaction is considered as being controlled by either the forward rate function or the thermodynamic function when the relevant ratio above is equal to 0.1. Thus kinetic control is indicated by $\rho \leq 0.1$ (from Equation 2.59), thermodynamic control by $\rho \geq 0.9$ (from Equation 2.60), and a combination of the two functions when $0.1 < \rho < 0.9$. From the comparison between these boundaries and the arbitrary boundaries set by Rolleston [135] shown in Fig. 2.11A, we can see that they are superior for defining reactions as near-equilibrium or far-from-equilibrium reactions as are equally far from 0 and 1 [16].

2.6.2 Rate capacity and mass-action contributions

Rate equations can also be expressed in a second way that separates the mass-action and rate capacity aspects. Thus the rate equation in Equation 2.54 for our mass-action reaction can be rewritten, in log form, as:

$$\ln v = \ln k_f + \ln\left(a^m b^n - \frac{c^p d^q}{K_{eq}}\right) \tag{2.61}$$

where the rate capacity term k_f is concentration independent, while the mass-action term $\left(a^m b^n - c^p d^q / K_{eq}\right)$ is concentration dependent. Like the $v_{(1-\rho)}$ function from the previous section, the mass-action function will be equal to 0 at equilibrium; however, unlike $v_{(1-\rho)}$, the mass-action function has different values for the same ρ depending on the substrate and product concentrations.

Rate equations of enzyme-catalysed reactions (which include the effects of catalysis and binding), can be rewritten in a similar manner[3]. For the reversible uni-uni reaction shown in Equation 2.7 the log form of a rate equation for enzyme-catalysed reactions would take the general form of:

$$\ln v = \ln(k_s e_0) + \ln\Theta + \ln\left(s - \frac{p}{K_{eq}}\right) \tag{2.62}$$

where $k_s = k_{cat}/K_s$ and denotes the enzyme's specificity constant for S, e_0 is the total enzyme concentration, and $k_s e_0$ (v_{cap}) is equivalent to k_f in Equation 2.61. A second difference is the inclusion of the binding term Θ, which expresses the effect of saturation. The final term is the mass-action term (v_{ma}). These terms can be plotted in a similar manner to Fig. 2.11 in order to visualise their contributions towards the total reaction rate.

As previously stated, metabolic regulation can be defined in terms of the augmentation or counteraction of the mass-action trend by the effect of an enzyme [25]. In order to determine in which direction the reaction is being pulled, elasticity coefficients can once more be used. The elasticity coefficients for the terms in Equation 2.62 are:

$$\varepsilon_s^v = \varepsilon_s^{v_{cap}} + \varepsilon_s^{\Theta} + \varepsilon_s^{v_{ma}} \tag{2.63}$$

where $\varepsilon_s^{v_{cap}}$ is the elasticity of the rate capacity term and is equal to zero, ε_s^{Θ} shows the contribution of enzyme binding towards the elasticity, and $\varepsilon_s^{v_{ma}}$ shows the contribution of mass-action towards the elasticity.

[3] Allosteric effectors are not included in any of the following examples as their effects are purely kinetic and therefore do not need to be separated from the mass-action effect.

Using the principles presented above, the reversible Michaelis-Menten equation for the uni-uni reaction can be expressed as:

$$v = \frac{V_f}{K_s} \times \frac{1}{1 + \frac{s}{K_s} + \frac{p}{K_p}} \times \left(s - \frac{p}{K_{eq}} \right) \tag{2.64}$$

where V_f is the limiting rate, and K_s and K_p are the respective Michaelis constants for S and P. The three terms (in log form) correspond to those of Equation 2.62. In Fig 2.12A, the individual terms are plotted against s in double-log space, as in Fig 2.11, thus showing their contribution towards v.

The slopes of the curves in Fig 2.12A are given by the elasticity coefficient expressions:

$$\begin{aligned} \varepsilon_s^v &= \frac{-\frac{s}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} + \frac{1}{1-\rho} \\ &= \varepsilon_s^{\Theta} + \varepsilon_s^{v_{ma}} \end{aligned} \tag{2.65}$$

which are obtained through partial differentiation of the log form of Equation 2.64 with respect to $\ln s$. Similarly, partial differentiation with respect to $\ln p$ leads to:

$$\begin{aligned} \varepsilon_p^v &= \frac{-\frac{p}{K_s}}{1 + \frac{s}{K_s} + \frac{p}{K_p}} + \frac{-\rho}{1-\rho} \\ &= \varepsilon_p^{\Theta} + \varepsilon_p^{v_{ma}} \end{aligned} \tag{2.66}$$

Depending on the values of s and p, the binding elasticity coefficients vary between 0 and -1, respectively, in these two equations. On the other hand, the mass-action elasticity coefficients vary between 1 and ∞, and 0 and $-\infty$ in Equations 2.65 and 2.66, respectively.

The elasticity coefficients in Equation 2.65 are plotted against s in double-log space in Fig 2.12B. The effect of substrate binding (ε_s^{Θ}) counteracts the mass-action effect ($\varepsilon_p^{v_{ma}}$) as the s increases and the reaction moves away from equilibrium. This can also be deduced from the signs of the terms in Equation 2.65. Conversely, while not shown here, product binding augments the mass-action effect. As noted in [25], this desensitisation towards substrate and sensitisation towards product is a natural consequence of a reaction being enzyme-catalysed. However, as also noted [25], while the position of equilibrium cannot be altered, the binding and catalytic properties of enzymes vary within the constraints of the Haldane relationship. Different binding and catalytic properties result in different magnitudes for the substrate and product binding effects and can potentially shift the position of the steady state in a system where these enzymes occur.

Finally, this method can also be applied in the analysis of reactions catalysed by enzymes with cooperative binding. In the rate equations of these reactions, the effect of binding is

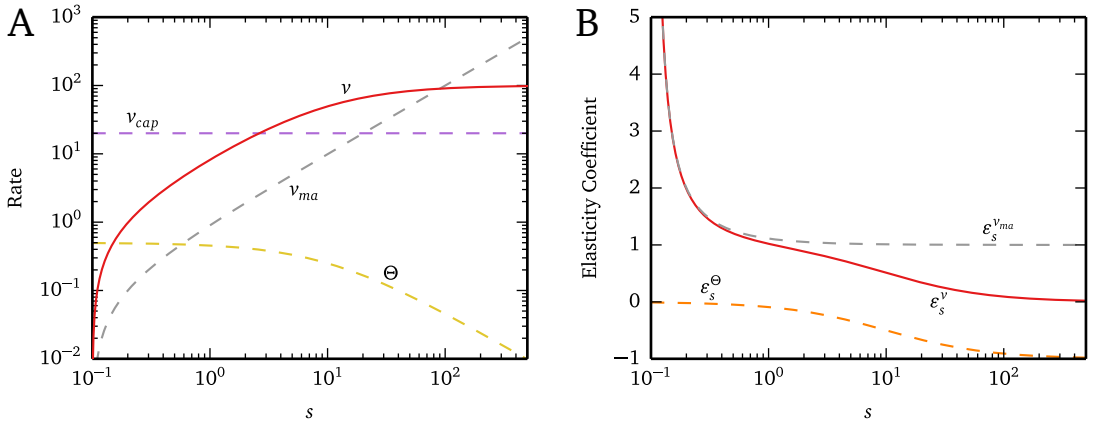


Figure 2.12: The rates and elasticity coefficients of a uni-uni reversible Michaelis-Menten reaction as a function of substrate concentration. In A the terms v_{ma}, v_{cap}, Θ and v are shown in double-log space against s. In B the elasticity coefficients corresponding to these terms are shown in double-log space against s. In both A and B $V_f = 100$, $K_{eq} = 10$, $K_s = 5$, $K_p = 1$ and $p = 1$. This figure was adapted from [16].

separated into two terms. The first term, Θ_1, represents the effect of single-subunit binding and is equivalent to the Θ binding term in the Michaelis-Menten rate equation. The second term, Θ_h, represents the effect of cooperative binding and results an elasticity coefficient profile that is distinct from that of the single subunit (see [16]).

2.7 Modelling and simulation software

Modelling of metabolic systems has a long history that precedes most of the theoretical frameworks discussed thus far. Early work was performed using analogue computers [137, 138] before later moving to digital computers [139, 140]. In either case, however, these methods were not widespread and only accessible to experts in computing technology. These early modelling efforts mostly involved performing time-course simulations of models, with Garfinkel and his co-workers being considered some of the pioneers in this field (for review see [141]). With the introduction of metabolic control analysis and the growth of the field of systems biology, however, a need for new types of software was established. Today, with the enormous technological advancements of the past few decades, this goal has been realised and hundreds of modelling tools have been developed thus far [142]. Below we will highlight some of the developments of the last 30 years within this field.

The earliest simulator that was available for computers other than mainframes was `METAMOD` [115]. It was developed using BASIC in 1986 for the 6502-based BBC Model B microcomputer under the Acorn operating system. This software package was also one of the

first to include functionality for performing MCA. METAMOD consisted of two programs that could be used to design and perform metabolic modelling experiments. METADEF was used for defining metabolic pathways which could then be loaded by METACAL. In turn METACAL could calculate the steady-state concentrations and fluxes of a model as well as the control and elasticity coefficients through a central difference method. METAMOD, together with a FOTRAN based program called MetaModel, was later used as the basis for developing package called MetaModel 2.0 [143]. MetaModel 2.0 was essentially an updated version of METAMOD which ran on IBM compatible computers under MS-DOS.

CONTROL [144] was released in the same year as MetaModel. However, unlike MetaModel, this software package was not developed as a general modelling tool, but rather focussed on performing metabolic control analysis. It was the first to employ the matrix-based approach developed by Reder [56] to calculate the values of control coefficient from elasticity coefficients. Additionally it could also provide symbolic expressions for the summation and connectivity properties. As mentioned in a previous section, other packages developed around the same time also attempted to provide symbolic expressions for control properties: an algorithm by Schulz [120] could provide symbolic control coefficient expressions for simple pathways, whereas MetaCon [22] was a more robust package which used a method based on the control-matrix equation to achieve the same goal for a wider variety of model structures.

Throughout the rest of the 1990s many other modelling packages were developed. SCAMP [114, 145] and Gepasi [117, 146] are two notable examples of popular modelling packages that had very different design philosophies while providing fairly comparable feature sets. SCAMP had a more traditional text based user interface, whereas Gepasi embraced the newer graphical user interface which was becoming increasingly prevalent in computing. While SCAMP originally calculated control coefficients using a finite difference and perturbation based method similar to that of MetaModel, was later updated to (like Gepasi) use Reder's [56] approach. Other software developed during this decade includes MIST [147] and DBsolve [148].

In the next decade there was a continued trend of development of modelling software. Here packages such as CellDesigner [149], ScrumPy [150] and VirtualCell [151, 152] were released. Notably SCAMP and Gepasi were respectively succeeded by the new tools called Jarnac [153] and Copasi [92]. These two new packages continued the legacy of their predecessors by retaining the same design philosophies, while providing a wider range of features. Copasi, for instance, includes functionality for building and editing models; performing time-course simulations, parameter scans, and stoichiometric analysis; as well as an array of data output and visualisation options. Jarnac also forms part of the suite of tools provided by the Systems Biology Workbench (SBW) [154]. This software package

aims to provide an extensible and modular software framework that allows for integration of different modelling tools via a variety of programming language bindings.

One contemporary of the above mentioned software tools that is of particular interest, is the Python Simulator for Cellular Systems or `PySCeS` [91]. This Python based open-source tool was developed after initial experimentation [155] with metabolic modelling using pure Python and SciPy [156]. Like similar software tools, `PySCeS` provides a variety of features, such as tools for performing time-course simulations, steady-state simulations, metabolic control analysis, and structural analysis. One of its most powerful features, however, stems from its design. Like `SCAMP`, `PySCeS` is a text based (or command line) tool without a GUI. It, however, makes full use of the Python programming environment and can be used as a library, in conjunction with any other Python libraries (see [157]), in order develop "scripts" for performing modelling experiments. In this sense `PySCeS` is more akin to `ScrumPy`, which is also a Python based modelling tool. While `ScrumPy` also provides functionality for dynamic modelling, it is more focused on structural modelling for large genome-scale models. Thus, while a feature-by-feature comparison with other modelling software might not indicate a clear advantage, `PySCeS` can be a very powerful tool as it allows for possibilities that are not necessarily provided in GUI environments or even foreseen by the original developers.

Some of the greatest advancements for metabolic modelling in the past 15 years have not been in the form of modelling software. One such advancement can be found in the development of the Systems Biology Markup Language (SBML) [158]. This language for defining systems biology models was developed in order to provide a unifying standard to ensure interoperability in a field that could otherwise suffer from fragmentation. SBML has been a great success when considering its widespread adoption; currently the SBML website lists more that 280 software packages that are SBML compatible [142].

Another great advancement has been the introduction of various online databases. Model databases, such as Biomodels [159] and JWS online [160], are of particular importance as they provide a central repository for storing and retrieving biological models. This development has significantly opened up the Systems Biology community as it encourages the sharing of data between groups from across the globe. These databases can therefore be considered to form part of the open science movement (see e.g. [161]). Additionally these databases curate incoming models in order to ensure that they yield the same results presented in their associated publications. In the case of JWS online, web-based model simulation features are also provided. Databases such as SABRIO-RK [162], BRENDA [163] or KEGG [164] do not provide models, but rather provide information about the components and reactions within biological systems and can thus potentially aid in the construction of models.

From this brief review of the landscape of metabolic modelling it should be clear that there

is already a great variety of options for investigating these models. Certainly some of these tools represent a duplication of effort, however, one must take into account the timespan of these developments. Since the publication of `METAMOD`, the technological landscape has changed immensely, and personal computers today are superior in every way to those of the 1980s. Additionally, the field of systems biology has grown through the acquisition of new knowledge and the development of new analysis frameworks. These advances warrant the development of faster and more efficient simulation tools, such as the recently developed `libRoadRunner` [165], and tools that implement new theoretical frameworks for metabolic analysis.

In the preceding sections of this chapter we have introduced various such metabolic analysis techniques that aim to reach a mechanistic understanding of the behaviour, control, and regulation of metabolic systems. Symbolic control analysis [17, 23, 119], generalised supply-demand analysis [15], and the thermodynamic/kinetic analysis framework of Rohwer and Hofmeyr [16], in particular, are squarely focussed on this goal. However, in spite of the fact that some of these methods have existed for many years, there is still a distinct lack of software solutions in which all of them are implemented. Furthermore, as each of these frameworks views metabolism from a different vantage point, they have the potential to complement each other in uncovering the properties of metabolic system. This potential has, as of yet, not been realised. Thus, in the following chapter we introduce our addition to the model analysis landscape in an attempt to provide a software solution that not only provides access to these analysis frameworks, but presents them in the form of a toolbox that facilitates their complementary use.

Chapter 3

PySCeSToolbox: Model Analysis Extensions for PySCeS

3.1 Introduction

One of the main goals of systems biology is to be able to explain the behaviour and characteristics of biological systems in terms of the properties and interactions of their components. The framework of metabolic control analysis [8, 9] is arguably one of the earliest examples of a method that endeavours to achieve this goal and has, since its inception more than 40 years ago, proven to be an invaluable tool in the study of metabolic systems (see e.g. [11, 36, 166]). Another indispensable tool is that of metabolic models, which have made it possible to perform *in silico* metabolic experiments with the power to predict and explain the behaviour of real metabolic systems. The last few decades have seen many other developments with the same ambitions in mind: high-throughput and "-omics" technologies have greatly increased the potential for data collection [7, 167], which has in turn led to larger and more complex metabolic models [18, 52]; and numerous techniques for investigating metabolic systems have been conceived [167–169]. Another source of progress within the field of systems biology comes from the advances made in terms of the processing power, computational techniques, and visualisation capabilities of computers. The recently published `libRoadRunner` SBML simulation and analysis library, for instance, can perform simulations orders of magnitude faster than similar software by making use of techniques such as just-in-time compilation [165]. These developments, in terms of new analysis techniques as well as software that capitalises on technological advances, provide new opportunities for understanding biological systems.

In the past our research group has been particularly interested in developing theoreti-

cal frameworks for investigating metabolic systems. Three such frameworks that we have contributed to are: metabolic supply-demand analysis (SDA) [14] and its generalised form (GSDA) [15], symbolic control analysis [22, 23], and a framework for investigating the thermodynamic and kinetic aspects of enzyme-catalysed reactions [16, 25]. These conceptual tools address a variety of questions regarding the behaviour, control, and regulation of metabolic systems; and each has, at least to some extent, been utilised in past metabolic studies [24, 26, 28–31]. However, their use in the investigation of metabolic models has been limited by a lack of robust and accessible software implementations. Additionally, a major shortcoming is that there has not yet been an easy way to apply these frameworks in conjunction with each other to answer inter-related questions regarding metabolic function. While implementations of GSDA and symbolic control analysis have been developed [15, 23], they can be regarded as prototypes rather than fully featured, interoperable software packages.

Here we therefore present `PySCeSToolbox`, a set of software tools aimed at overcoming these limitations. `PySCeSToolbox` is developed as an extension to the Python Simulator for Cellular Systems (`PySCeS`) with the goal of providing access to the above-mentioned theoretical frameworks for use in model analysis. It also allows the user to utilise these frameworks in a complementary fashion with relative ease. The toolbox includes three main tools: `RateChar`, which is used to perform GSDA [15]; `SymCa` for performing symbolic metabolic control analysis and control-pattern analysis [17, 23, 119]; and `ThermoKin`, a tool for applying the kinetic/thermodynamic investigation framework of Rohwer and Hofmeyr [16]. While versions of `RateChar` and `SymCa` have been developed in the past, these new versions have been completely rewritten from scratch, and all three of the new tools have been designed with a consistent API, shared functionality and data structures, and with ease-of-use in mind. Moreover, `PySCeSToolbox` has been developed to integrate with the scientific Python ecosystem.

3.2 Design and architecture

One of the main goals of `PySCeSToolbox` is to provide a powerful set of analysis tools that is accessible to users with varying levels of technical skill. A second important goal is that the software should be compatible with a variety of platforms and fit in with the current ecosystem of the user.

`PySCeSToolbox` is developed in Python and is designed as a library of extensions to `PySCeS`. Besides the fact that using Python greatly simplifies the task of integrating these software, this language and design was chosen for two main reasons. Firstly, designing `PySCeSToolbox` as a library allows for modelling experiments to be prepared as programs,

or scripts, in which the steps of an experiment are typed into a source code file as a sequence of commands that are subsequently executed consecutively. While a live code interpreter can also be used, where each step is executed individually after being entered, the scripting paradigm facilitates reproducibility and repeatability, and allows for more complex experiments to be designed. As will be discussed below, using Python also allows for a hybrid approach via the `Jupyter Notebook`. Secondly, the choice of Python provides users access to the multitude of scientific and general purpose libraries that have made this language a standard tool for scientific computing [157]. While some similar functionality is provided by other scripting languages, such as Julia [170] and Perl (with the perl data language) [171], none provide the same level of maturity and variety of third-party libraries as Python. Ultimately, these design choices allow for immense flexibility, as both simple and enormously complex tasks (not necessarily envisioned by the developers) can be performed efficiently and reproducibly with relative ease.

Another important component of the computing platform is the `Jupyter Notebook` [172], which acts as the interface between the user and Python via the `IPython` kernel [173]. This software package is a web application, which allows users to create notebook documents that contain not only code, but also markdown formatted text, LaTeX formatted equations, and data visualisations. In certain aspects the `Jupyter Notebook` paradigm is similar to that found in computer algebra systems such as `Mathematica` [90] and `SageMath` [88] and should feel familiar to users of these packages. On the front-end, `Jupyter` utilises a variety of web technologies to provide an easy-to-use interactive coding interface where individual input cells are followed by appropriate outputs. It also provides widget support where code and visualisations can be manipulated interactively through standard graphical user interface (GUI) elements without the need for in-depth programming knowledge. This component, while not essential for using the core features of `PySCeSToolbox`, therefore greatly streamlines the experience of writing code and generating results.

In addition to making use of many built-in Python and `Jupyter Notebook` features, a variety of third-party packages from the scientific Python ecosystem are used to provide `PySCeSToolbox`'s functionality. All plotting is performed via an interface to the `Matplotlib` 2D plotting library [174] and plots can be customised using standard `Matplotlib` functionality. Symbolic computation and expression manipulation is predominantly performed using the `SymPy` library [175]. Additional symbolic computation is provided by the `Maxima` computer algebra system [176]; however, we plan on removing this dependency in the future. Metabolic pathway figures are drawn using a combination of features from the `D3.js` Javascript visualisation library [177] and the `NetworkX` Python package for studying graphs and networks [178]. Finally, a variety of miscellaneous features are provided by the `NumPy`

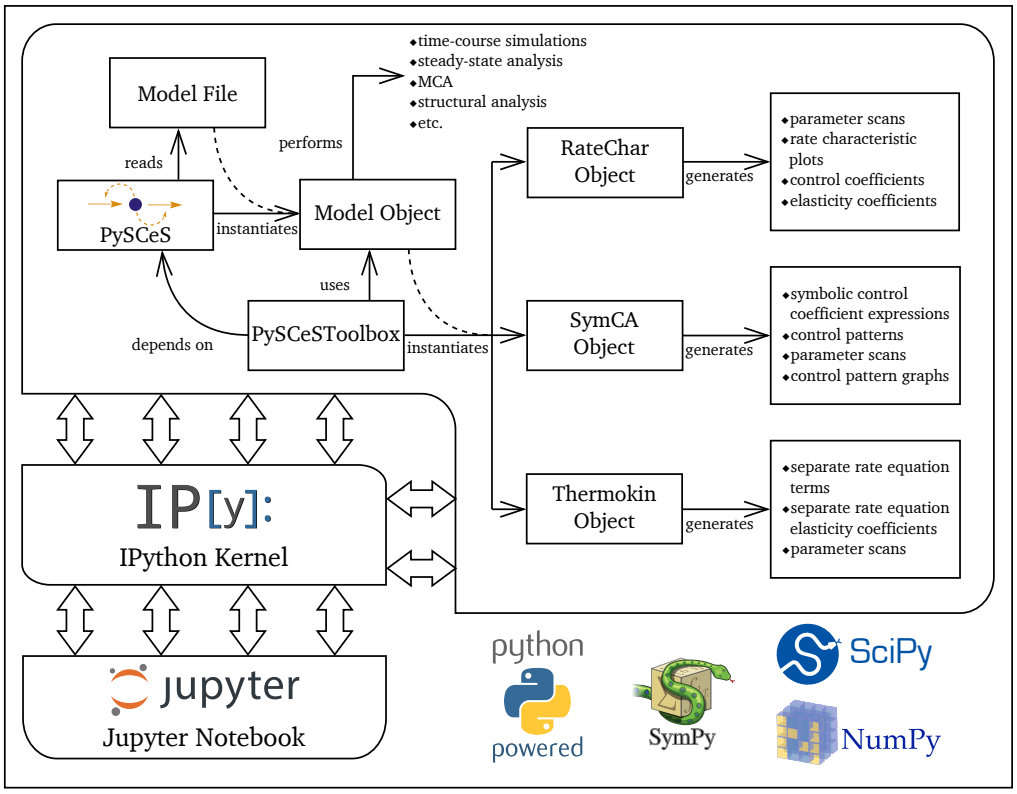


Figure 3.1: `PySCeSToolbox` architecture and workflow. Pysces reads a model file and instantiates a model object, which is in turn used by `PySCeSToolbox` to instantiate one of the analysis tool objects. The three analysis tools each generate a set of results according to, when appropriate, user-defined parameters. The recommended usage involves running all these processes within an IPython kernel, with which the user interacts via the Jupyter Notebook. The bottom-right corner shows some of the main technologies used by `PySCeSToolbox`.

N-dimensional array library [179] and the `SciPy` library for scientific computing [156].

Fig. 3.1 provides a simplified scheme of how the above-mentioned components fit together and also shows a basic workflow. Finally, while `PySCeSToolbox` was developed primarily using the Linux operating system, it is cross-platform compatible and has been tested to work on both Windows (tested on Windows 7, 8, and 10) and Mac OSX.

3.3 General features

As previously mentioned, the bulk of the functionality provided by `PySCeSToolbox` stems from its three main components: `RateChar`, `SymCa`, and `ThermoKin`; where each corresponds to a different analysis framework. These tools do, however, share a variety of common features and functionality.

Most obviously, much of the design and features stem from the base package PySCeS. Firstly, a similar workflow is followed where analysis tool objects are instantiated from PySCeS model objects. Each tool is subsequently responsible for generating, storing, visualising, and exporting results (see Fig. 3.1). Secondly, for the sake of consistency, PySCeSToolbox inherits its variable naming scheme from PySCeS[1]. Finally, some of the results generated by PySCeS can be visualised using the functionality provided by PySCeSToolbox.

In addition to its dependence on PySCeS, much of PySCeSToolbox's strength lies in its utilisation of the visualisation functionality of the Jupyter Notebook. One of the most useful features of our software is thus the interactive plotting interface provided by each of our analysis tools. Here generated results can be visualised as an appropriate 2D-plot. As shown in Fig 3.2, these plots can subsequently be manipulated in terms of the displayed results (as well as technical details such as axes-limits and scale) using a set of GUI elements in addition to the standard programmatic commands. Additionally, visualisation features include the ability to display data structures that contain results as human readable tables, and the ability to display symbolic expressions as LaTeX formatted equations; all within the notebook.

PySCeSToolbox allows the user to save work in a variety of ways. Figures can be exported in any of the formats supported by Matplotlib. Data from parameter scans and all of the tools discussed in the following sections can also exported for use in other software packages as comma separated value (.csv) files. Moreover, each of RateChar, SymCa, and ThermoKin have the ability to save the data of the current analysis session for later use.

Finally, PySCeSToolbox also provides some functionality for displaying metabolic pathway schemes in the notebook, as demonstrated in Fig 3.4. This functionality is predominantly used within SymCa (as will be discussed in Section 3.5), but can be accessed as a stand-alone feature. While pathway layouts can be saved, no automatic layout algorithms are provided, and therefore pathway elements have to be arranged manually.

3.4 RateChar

The RateChar tool is used to perform generalised supply-demand analysis on models of metabolic pathways. Below follows a short description of this method and its uses, the procedure for performing GSDA, and an example of results produced by RateChar.

Supply-demand analysis is a quantitative framework for investigating the control, regulation and behaviour or metabolic pathways [14]. Here metabolic pathways are viewed as

[1]We refer the reader to the documentation of PySCeS and PySCeSToolbox, located respectively at http://pythonhosted.org/PySCeS/ and http://pyscestoolbox.readthedocs.io.

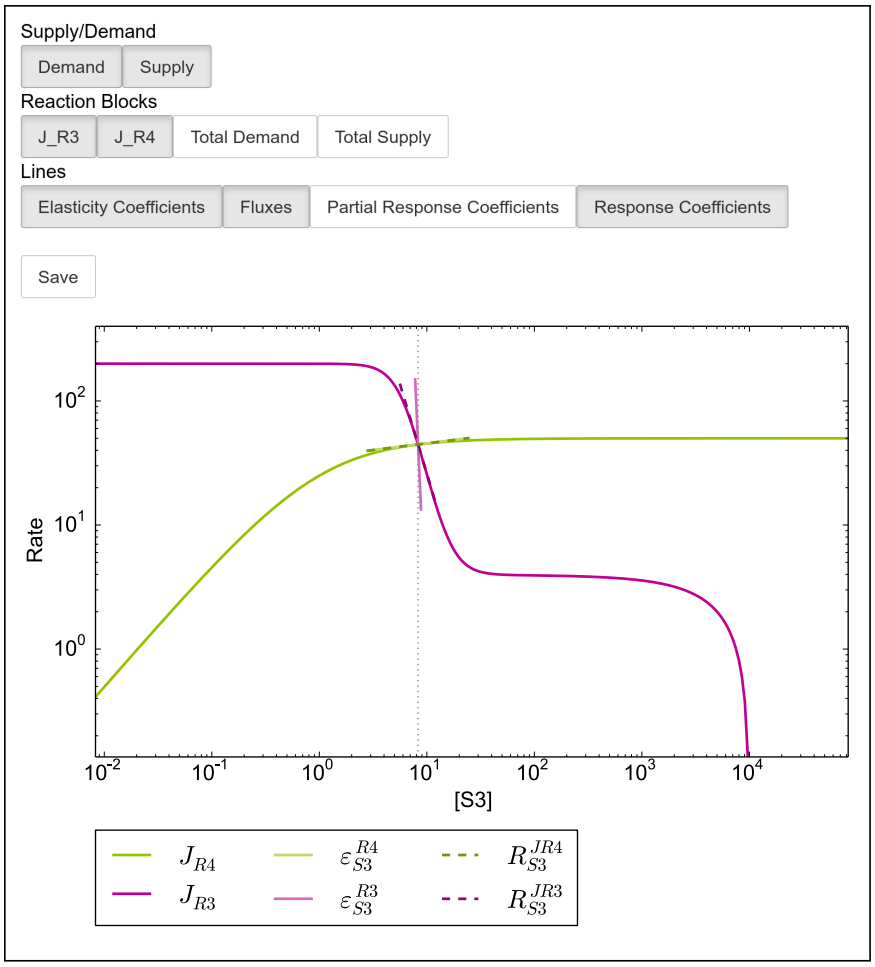


Figure 3.2: An example of a 2D-plot generated by `PySCeSToolbox`. This screenshot demonstrates `PySCeSToolbox`'s GUI elements that can be used to modify the plots it produces from within a Jupyter Notebook session. Here a rate characteristic plot for the supply and demand blocks of S_3 in Fig. 3.3 was produced using `RateChar`. The fluxes of these blocks are shown as functions of the intermediate concentration in double-log space. The response coefficients of these reaction blocks towards S_3 are indicated as dashed tangents to the curves at the steady state (shown as a vertical dotted line), whereas elasticity coefficients are indicated as solid lines intersecting the steady-state point. Here the legend reflects the names of the reactions and metabolites as they are defined in the model. As indicated by the various buttons above the plot, the specific configuration of data shown can be altered. Lines pertaining to a specific category of results, such as supply, demand or a specific reaction block, can be enabled or disabled as a group. Similar functionality is included for `SymCa` and `ThermoKin`, and arbitrary PySCeS parameter scans and time-course simulations can also be plotted using the same interface.

molecular economies in which certain processes have evolved to fulfil the role of a "supply" block that produces a certain key intermediate, and a "demand" block that subsequently consumes this intermediate. A great emphasis is therefore placed on investigating metabolic

pathways within the context of their functional organisation. According to the original authors of this framework [14, 37], the demand for a certain intermediate can often be expected to be responsible for controlling flux through a pathway, whereas the supply of the intermediate could be expected to determine its degree of homoeostasis.

In generalised supply-demand analysis, the need for choosing a linking metabolite between reaction blocks is avoided by considering each intermediate, in turn, as a linking metabolite [15]. In this way the selection bias by a researcher is avoided, and detection of the point of functional differentiation (if such a point exists) comes about naturally from the analysis. GSDA therefore also identifies potential sites of regulation and regulatory metabolites. Additionally, in the case where there are multiple routes of interaction between an intermediate and a reaction block, GSDA can determine the quantitative relative contribution of each of these routes towards the flux response of the block with respect to the intermediate.

The two main tools of GSDA are rate characteristic plots [25] and metabolic control analysis [8, 9]. A rate characteristic plot is drawn for each intermediate in a system and shows the fluxes of its supply and demand blocks as a function of its concentration, as demonstrated in Fig. 3.2. Response and elasticity coefficients are also indicated on the rate characteristic plot. While GSDA can theoretically be applied to *in vivo* metabolic systems, as has been done with SDA (e.g. [28, 29]), practical considerations limits its use in experimental systems. GSDA is, however, ideally suited for *in silico* modelling experiments.

The `RateChar` tool automates the process of performing GSDA on models of metabolic pathways. This procedure consists of a series of steps which are performed for each intermediate in a pathway. These steps are as follows:

1. The steady-state concentration of the intermediate is determined.

2. The concentration of the intermediate is fixed at its steady-state value, turning it into a parameter of the system.

3. Metabolic control analysis is performed and the total and partial response coefficients of the reaction blocks with respect to the intermediate, as well as the elasticity coefficients of the reactions interacting with the intermediate, are determined.

4. A parameter scan of the fixed intermediate is performed where its concentration is varied over a range of values, which (by default) spans four orders of magnitude with the reference steady state at the centre. A new steady state is determined for each of the points in this range.

5. The reaction block fluxes are plotted as functions of the fixed intermediate on a rate characteristic plot together with the metabolic control analysis results. Each reaction

that consumes or produces a metabolite is considered by `RateChar` as the starting point or end point of a reaction block.

Here steady-state and metabolic control analysis calculations are performed using the functionality provided by `PySCeS`. Once this process is set in motion, each of these steps is automatically carried out, and user intervention is only required when visualising results.

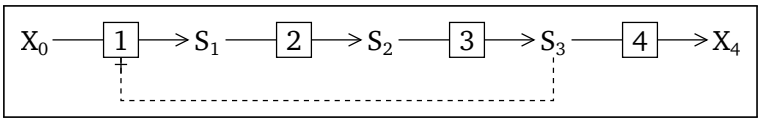


Figure 3.3: A 4-step pathway with allosteric inhibition. Here X_0 is converted to X_4 by reactions 1–4 via the intermediates S_1–S_3. Reactions 2–4 follow Michaelis-Menten kinetics, with reactions 2 and 3 being reversible and reaction 4 being irreversible. Reaction 1 follows reversible Hill kinetics and is allosterically inhibited by S_3.

While these steps can be executed using a set of default options, much of the process can be customised by the user: a subset of metabolites for which parameter scans should be performed can be selected, the scan range can be manually-defined, and the specific steady-state solver to be used can be selected (see [91]). Finally, as previously mentioned, rate characteristic plots can be customised using a GUI and saved as any of the graphics formats supported by `Matplotlib`.

The time taken to perform a parameter scan using `RateChar` depends on the steady-state solver used, the size of the model, the number of points to scan as well as the difficulty of determining a steady-state solution for any given point. For example, performing a parameter scan using the default settings, where the steady state is solved for 256 points per variable intermediate in a single-threaded process, takes an average of 13.7 seconds per intermediate (or pair of intermediates in a two-member moiety-conserved cycle) on a dual core intel Core i5 mobile processor with a maximum core frequency of 3.1 GHz based on a model of pyruvate metabolism in *Escherichia coli* [31, 86]. However, the fastest scan took just under an average of five seconds to complete, whereas the slowest took just over 35 seconds.

By using `RateChar` to perform GSDA on a model of the 4-step pathway shown in Fig. 3.3, rate characteristic plots are generated for each of the intermediates S_1–S_4. Fig. 3.2 shows the rate characteristic plot for S_3. This system can be regarded as functionally differentiated at S_3; the demand block controls the pathway flux, while the supply block determines the degree of S_3 homoeostasis. This is deduced from the fact that the gradient of the demand block curve is nearly 0 at the steady state, whereas the supply block curve gradient is steep and negative, as expressed by $|R^{J_3}_{s_3}/R^{J_4}_{s_3}| > 1$ (see [14]). If S_3 demand is decreased so that its curve intersects the supply curve at $S_3 \approx 10^4$, then flux control would still reside with the S_3 demand block,

albeit at S_3 concentrations near chemical equilibrium of the S_3 supply block. However, by increasing the S_3 demand so that its curve intersects the supply curve in the region where $S_3 \lesssim 2 \times 10^0$, flux control would shift to the supply block as here $|R^{J_3}_{s_3}/R^{J_4}_{s_3}| < 1$.

3.5 SymCa

`SymCa` is used to perform symbolic control analysis on metabolic models. As in the section above, we will describe this method and its application through `SymCa` together with an example.

Metabolic control analysis was originally developed by in the early 1970s by two independent research groups [8, 9] and provides a framework for the quantitative analysis of the control of metabolic systems. Since its inception it has largely debunked the idea of an obligatory rate limiting step and is most probably one of the earliest "systems biology frameworks" (see Fell [12] for an introductory text). In MCA the sensitivity of a system variable (y), such as a flux or steady-state concentration, towards a reaction step (i) is described by a control coefficient (C^y_i), whereas the sensitivity of a rate of such a step (v_i) towards a variable or parameter (x) in isolation is described by an elasticity coefficient ($\varepsilon^{v_i}_x$). A great strength of MCA is that the values of the control coefficients of a system can be determined from the values of the elasticity coefficients of its enzyme components through the application of the summation and connectivity theorems of MCA [8], which allow the control coefficients to be expressed analytically in terms of elasticity coefficients.

A number of different matrix based formalisations of MCA exist (e.g. [19–21, 56, 57, 98–110]) which simplify the derivation of control coefficients from elasticity coefficients. Here we make use of the form by Hofmeyr [57] in which the summation and connectivity theorems are combined into a single expression $\mathbf{C}^i\mathbf{E} = \mathbf{I}_n$, where $\mathbf{C}^i$ is a matrix of the independent control coefficients, $\mathbf{E}$ is matrix of structural data and elasticity coefficients, and $\mathbf{I}_n$ is an identity matrix (see also Section 2.4.1). Using this control matrix equation, the control coefficients can be determined by inversion of the $\mathbf{E}$ matrix.

In symbolic control analysis the inversion of the $\mathbf{E}$ matrix is performed algebraically, rather than numerically, thus producing symbolic control coefficient expressions. Each control coefficient of a pathway is expressed as a fraction with a unique numerator, and a denominator that is shared among all the control coefficients of the pathway. Both the denominator and the numerator consist of multiple terms that are expressed in a standard form [23]: each term is the product of multiple elasticity coefficients; if a branch point is present in a pathway, the terms are multiplied by fluxes; and, if a term contains an elasticity coefficient with respect to a species that forms part of a moiety-conserved cycle, that term will be divided by the

concentration of the species. Finally, the common denominator is given by the determinant of **E**, and is the sum of all the control coefficient numerators of the pathway.

These symbolic control coefficient expressions have certain advantages over the mere numerical values:

- By investigating the individual control expression terms using a technique called control-pattern analysis [17, 119], the control structure of a metabolic pathway can be dissected.
- Control coefficient expressions depend only on pathway stoichiometry and the knowledge of regulatory (allosteric) interactions, and therefore allow for some general conclusions regarding control structure without the need for knowledge of the complete pathway kinetics.
- They have some predictive power regarding the effect of perturbations on the control of a system. This is especially true if certain assumptions are made regarding the properties of the reactions of the system at the steady state (e.g., assuming that certain enzymes are saturated, which results in zero elasticity coefficients, or that certain reactions have first-order rates, which results in elasticity coefficients of one), which allows for control coefficient expressions to be simplified.

In essence, symbolic control analysis is an explanatory tool for understanding and exploring the control properties of metabolic systems.

As with `RateChar`, `SymCa` fully automates the process of generating symbolic control coefficient expressions, as well as the subsequent analysis of these expressions. `SymCa` generates control coefficient expressions through the following steps:

1. A symbolic **E** matrix is constructed using the numeric matrices provided by the `PySCeS` model as a basis (see [57] for a description of the procedure used).
2. **E** is algebraically inverted using the cofactor method.
3. The resulting control coefficient expressions are simplified through factorisation.
4. The simplified expressions are rearranged into the standard form.

Here `SymPy` is used to perform the symbolic matrix inversion and to manipulate the resulting expressions. Due to its significant speed advantage, `Maxima` rather than `SymPy` is used to factorise the expressions generated through inversion.

The time taken to perform a symbolic inversion of the **E** matrix and the subsequent simplification of the resulting expressions depends on the size and complexity of the model, with

the number of interactions between intermediates and reactions causing an increase in the run time. Thus two models with the same number of reactions can have significantly different inversion and simplification times if their overall complexities differ in terms of the number of connections between their reactions. Symbolic inversion and expression simplification by `SymCa` is also much slower on any given model than a full model parameter scan by `RateChar`. For instance, this process takes over 37 minutes to complete on the previously mentioned *E. coli* pyruvate metabolism model using the same dual core intel core i5 mobile processor configuration, compared to the two-minute average run time of `RateChar`.

After the control coefficient expressions are generated, `SymCa` provides a variety of functionality for further analysis. In addition to providing numeric values for the control coefficients, the values of each term of these control coefficients (known as control patterns) are also calculated. These control patterns represent the chains of local effects in a pathway that describe the different paths through which perturbations of reaction steps can propagate through a system [17, 119], and can be visually represented on a metabolic pathway scheme (Fig. 3.4). Values of the control coefficients and control patterns are dynamically updated whenever the numeric metabolic control analysis function of `PySCeS` is performed on the model in question, thus allowing for the effect of changes in model parameters on the values of these expressions to be investigated.

Parameter scans can also be performed for the control coefficient expressions, which results in 2D-plots (similar to Fig. 3.2) of either the values of the control patterns of a control coefficient as a function of the parameter, or the percentage contribution of the control patterns towards the control coefficient. These percentages are calculated as the absolute value of a control pattern relative to the sum of the absolute values of all the control patterns of a control coefficient. This method is used instead of conventional percentages to account for cases where both negative and positive control patterns occur in the same control coefficient, e.g., for a control coefficient with two control patterns that are equal in magnitude, but opposite in sign, each is regarded as being equally responsible for (contributing 50% to) the zero control coefficient value.

For the example pathway in Fig. 3.3, the flux control coefficients for each of the four steps,

as generated by `SymCa`, are as follows:

$$C_1^J = \left(\varepsilon_{s_1}^{\nu_2}\varepsilon_{s_2}^{\nu_3}\varepsilon_{s_3}^{\nu_4}\right)/\Sigma \tag{3.1a}$$

$$C_2^J = \left(-\varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_3}\varepsilon_{s_3}^{\nu_4}\right)/\Sigma \tag{3.1b}$$

$$C_3^J = \left(\varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_2}\varepsilon_{s_3}^{\nu_4}\right)/\Sigma \tag{3.1c}$$

$$C_4^J = \left(-\varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_2}\varepsilon_{s_3}^{\nu_3} - \varepsilon_{s_3}^{\nu_1}\varepsilon_{s_1}^{\nu_2}\varepsilon_{s_2}^{\nu_3}\right)/\Sigma \tag{3.1d}$$

$$\text{with } \Sigma = -\varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_2}\varepsilon_{s_3}^{\nu_3} + \varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_2}\varepsilon_{s_3}^{\nu_4} - \varepsilon_{s_1}^{\nu_1}\varepsilon_{s_2}^{\nu_3}\varepsilon_{s_3}^{\nu_4} - \varepsilon_{s_3}^{\nu_1}\varepsilon_{s_1}^{\nu_2}\varepsilon_{s_2}^{\nu_3} + \varepsilon_{s_1}^{\nu_2}\varepsilon_{s_2}^{\nu_3}\varepsilon_{s_3}^{\nu_4} \tag{3.1e}$$

In these expressions Σ represents the common denominator of the control coefficients. Each elasticity coefficient represents the propagation of the perturbation though a metabolite-enzyme pair. In a linear pathway such as in this example, each control pattern will have a positive effect assuming 'normal' kinetics (i.e. substrates increase and products decrease the reaction rate). Thus the function of the positive or negative signs in these terms is to ensure that the overall sign of the control patterns, when taking the values of the individual elasticity coefficients into account, matches that of the real effect of the control pattern. In the case of steps 1–3, each control coefficient has a single control pattern, which represents the chain of effects that are propagated through the main metabolic branch by interactions of substrates and products with their enzymes. Reaction 4, on the other hand, has an additional control pattern, which represents the chain of effects which propagates through the feedback loop form S_3 to reaction 1. The control patterns of the main branch and the feedback loop are shown in Fig. 3.4A and B for the steady state indicated in Fig. 3.2. At this reference steady state, the feedback loop chain is completely responsible for determining C_4^J, whereas the main branch chain plays no role. As will be shown in the next section and in Fig. 3.5, lowering the demand for S_3 decreases the value of $\varepsilon_{s_3}^{\nu_1}$ to zero and drastically increases the value of $\varepsilon_{s_1}^{\nu_1}$. While this change does not shift control away from reaction 4 (as discussed in the previous section), it does affect the overall values of the two control patterns of C_4^J, thus shifting the source of control to the main branch away from the feedback loop.

3.6 ThermoKin

The third major tool provided by `PySCeSToolbox` is `ThermoKin`. This software tool is used to assess the kinetic and thermodynamic aspects of enzyme catalysed reactions.

According to Reich and Sel'kov [38] (see also Hofmeyr [25]), metabolic regulation can be defined as *the alteration of the reaction properties to augment or counteract the mass-action trend in a network of reactions*. In addition to the effect of the evolved stoichiometry of a metabolic network, this function is largely fulfilled by the enzymes catalysing the reactions

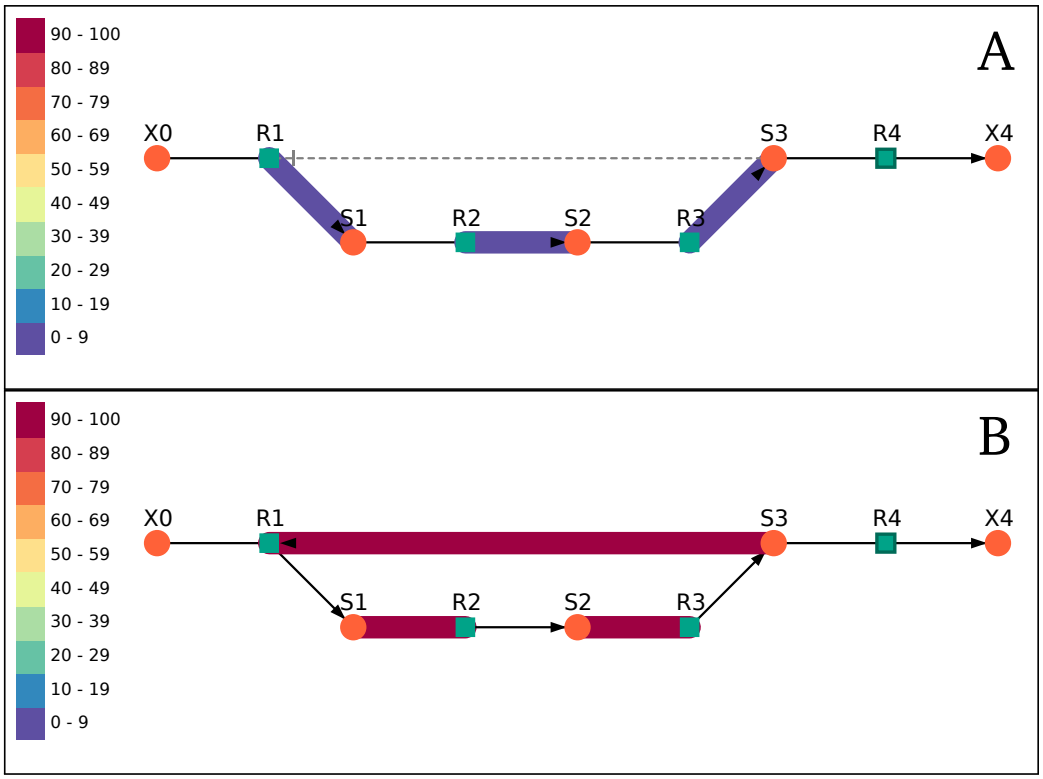


Figure 3.4: Examples of metabolic pathway schemes generated by `PySCeSToolbox`. These two schemes were produced using `SymCa` and shows the control patterns of C_4^J in the pathway of Figure 3.3. Circles represent species, while squares represent reactions, and the layout of these nodes was defined by hand. **A** corresponds to the first numerator term of Equation 3.1d and represents the propagation of a perturbation of reaction 4 via the main branch of the pathway at the steady state. **B** corresponds to the second numerator term and represents the propagation of the perturbation of reaction 4 via the feedback loop at the steady state. The colour of the highlighted control pattern indicates its percentage contribution towards the total control coefficient according to the colour bar on the left.

of the system. The regulatory potential of these enzymes is dictated by the distance of their reactions from equilibrium, as signified by the disequilibrium ratio Γ/K_{eq} (or ρ); far-from-equilibrium reactions are more susceptible to the effects of enzyme binding and catalysis than near-equilibrium reactions where the mass-action effect dominates.

Traditionally, reactions have been classified as either far-from-equilibrium or near-equilibrium according to the definition of Rolleston [135], who asserted that reactions with $\rho \leq 0.05$ are far-from-equilibrium and those with $\rho \geq 0.2$ are near-equilibrium. Relatively recently, this definition has been disputed by Rohwer and Hofmeyr [16], where $\rho \leq 0.1$ and $\rho \geq 0.9$ are presented as more appropriate delimiters for defining reactions as predominantly kinetically controlled (far-from-equilibrium) or thermodynamically controlled (near-

equilibrium).

Additionally, these authors provide a framework for separating the contributions made by enzyme binding and catalysis towards the rate of a reaction from that of mass-action. Here rate equations describing reversible enzyme catalysed reactions are cast in a form which clearly separates these components, e.g., the rate equation for the enzyme catalysed reaction S $\rightleftharpoons$ P can be written as:

$$v = k_s e_0 \times \Theta \times \left(s - \frac{p}{K_{eq}} \right) \tag{3.2}$$

where $k_s = k_{cat}/K_m$ and is the specificity constant of the enzyme for S, which, together with the total enzyme concentration e_0, denotes the rate capacity term (v_{cap}) of the enzyme. The second term (Θ) is the binding term and quantifies the effect of enzyme saturation on the reaction rate. The final term is the mass-action term (v_{ma}) of the reaction. While the expressions for v_{cap} and Θ are dependent on the kinetics of a reaction, v_{ma} will have the same general form regardless of the reaction, and only depends on the number of substrates and products of the reaction, together with their stoichiometric coefficients.

By partially differentiating the terms[2] of the logarithmic form of the rate equation above with respect to $\ln s$, the elasticity coefficients of the individual terms can be obtained:

$$\varepsilon_s^v = \varepsilon_s^{v_{cap}} + \varepsilon_s^{\Theta} + \varepsilon_s^{v_{ma}} \tag{3.3}$$

where $\varepsilon_s^{v_{cap}}$ is always 0, and ε_s^{Θ} and $\varepsilon_s^{v_{ma}}$ quantify the extent to which the substrate affects the reaction rate through enzyme binding and through the mass-action effect, respectively. Similarly, elasticity coefficients can be derived for the product. This analysis can be extended to include the effect of cooperative binding, and other authors have suggested slightly different ways of splitting rate equations to highlight different effects [129, 180].

`ThermoKin` provides a variety of features that facilitates the application of this framework in model analysis. Most importantly, it provides functionality for determining the distance of reactions from equilibrium in a steady-state model, and it allows a user to investigate the effects of the different rate equation terms and their corresponding elasticity coefficients as described above.

First, however, the rate equations of a model have to be separated into their different terms. Here a user has the option of manually separating each of the pertinent rate equations into a set of terms that are then read in from file by `ThermoKin`. While this option is the most labour intensive, it is also the most flexible and reliable, as the user can split the rate equations

[2] In the linear form of the rate equation (Equation 3.2), these terms are actually factors, but for the sake of simplicity we refer to them as terms as per Rohwer and Hofmeyr [16].

in any arbitrary manner. A second option is provided where `ThermoKin` will attempt to automatically separate the terms of the rate equations of a model into a mass-action term (v_{ma}) and a combined rate capacity and binding term ($v_{cap}\Theta$). In this case, delineating between the contribution of v_{cap} and Θ towards the total rate is not possible, however, the elasticity coefficient of this combined term is exactly the same as that of the binding term alone since $\varepsilon_s^{v_{cap}} = 0$. Regardless of how the rate equations are separated, `ThermoKin` will additionally attempt to automatically determine the expression for ρ of each reversible reaction.

The algorithm for separating a rate equation into v_{ma} and $v_{cap}\Theta$, and for determining the expression of ρ, depends only on the rate equation expression and knowledge of the substrates and products of the reaction. This algorithm does not rely on any specific naming scheme for species or parameters within the model, but rather depends on the form of the rate equation expression itself and should be compatible with any standard reversible rate equations (such as reversible Michaelis-Menten, reversible Hill, and reversible Monod-Wyman-Changeux kinetics). Unfortunately, however, the algorithm is not able to distinguish between the effect of cooperative and single-subunit binding (see [16]), and these two effects are lumped together in the kinetic term. The procedure commences with the assumption that a reversible rate equation will have one positive term, which contains the substrates, and one negative term, which contains the products. At its core it aims to identify the symbol that represents K_{eq} in each reversible rate equation of a model, as well as to construct the expression for Γ of each model reaction. Using functionality provided by `SymPy`, the following steps are performed:

1. The rate equations of a model are expanded and split into their terms.

2. Rate equations that have a single term are classified as irreversible, whereas those with two terms, where one is negative and the other positive, are classified as reversible. All subsequent steps are applied only to reversible reactions.

3. The expression for ρ of each reaction is obtained by dividing the absolute value of its negative term by its positive term, thus removing all other symbols.

4. The symbol representing K_{eq}, together with the expression of Γ, are determined from ρ and the substrates and products for each reaction.

5. The K_{eq} symbols, Γ expressions, products, and substrates of each reaction are used to construct the v_{ma} expressions.

6. Finally the $v_{cap}\Theta$ expressions are obtained by dividing each original rate equation by its corresponding v_{ma} expression.

After the separated rate equation terms are loaded by `ThermoKin`, they are available as symbolic expressions in a similar manner as those generated for control coefficients by `SymCa`. Additionally, their elasticity coefficients are automatically determined through symbolic partial differentiation with respect to their substrates, products, allosteric effectors, and parameters. Similar to `SymCa`, the values of the various terms and their elasticity coefficients are dynamically calculated whenever the steady state of the model is determined, thus allowing for their investigation under different sets of conditions.

Finally, parameter scans can be performed for either the rate terms or for their elasticity coefficients, allowing the user to assess the effect of change in a model parameter on the effects of the kinetic and thermodynamic aspects of a reaction on its rate. Here 2D-plots of the results can be generated similar to those of `RateChar` and `SymCa`.

Using the automatic rate equation splitting functionality of `ThermoKin`, the reversible Hill equation of reaction 1 in the example pathway of Fig. 3.3 is split into the combined binding and rate capacity, and the mass-action terms as follows:

$$v_1 = \underbrace{\frac{\frac{V_{f1}}{x_{0_{0.5}}} \times (\sigma+\pi)^{h-1}}{(\sigma+\pi)^h + \frac{1+\xi^h}{1+\alpha\xi^h}}}_{v_{cap}\times\Theta} \times \underbrace{\left(x_0 - \frac{s_1}{K_{eq^1}}\right)}_{v_{ma}} \tag{3.4}$$

where V_{f1} is the limiting forward rate. Additionally σ, π, and ξ respectively represent $x_0/x_{0_{0.5}}$, $s_1/s_{1_{0.5}}$, and $s_3/s_{3_{0.5}}$. Here $x_{0_{0.5}}$, $s_{1_{0.5}}$, and $s_{3_{0.5}}$ are the half saturation constants of X_0, S_1, and S_3.

Performing a parameter scan of the limiting rate of reaction 4 (V_{f4}), which represents a modulation of this reaction's total enzyme concentration, yields the elasticity coefficient results for reaction 1 shown in Fig. 3.5. This figure shows that the rate of reaction 1 is controlled by the effects of substrate binding ($\varepsilon_{x_0}^{v_{1\Theta}}$) and inhibitor binding ($\varepsilon_{s_3}^{v_{1\Theta}}$) in the reference steady state shown in Fig. 3.2. At lower V_{f4} values, however, the effect of inhibitor binding decreases to 0, whereas the mass-action effect ($\varepsilon_{x_0}^{v_{1ma}}$ and $\varepsilon_{s_1}^{v_{1ma}}$) takes over control of this reaction's rate. This change in the value of V_{f4} essentially represents a modulation of the demand for S_3 and is thus equivalent to "shifting" the curve representing this demand in Fig. 3.2 up and down. Therefore, while the control of flux does not shift away from reaction 4 at lower S_3 demand, as discussed in Section 3.4, the mass-action effect is responsible for control at low demand, whereas allosteric inhibition is responsible for control at higher demand as mentioned in the previous section.

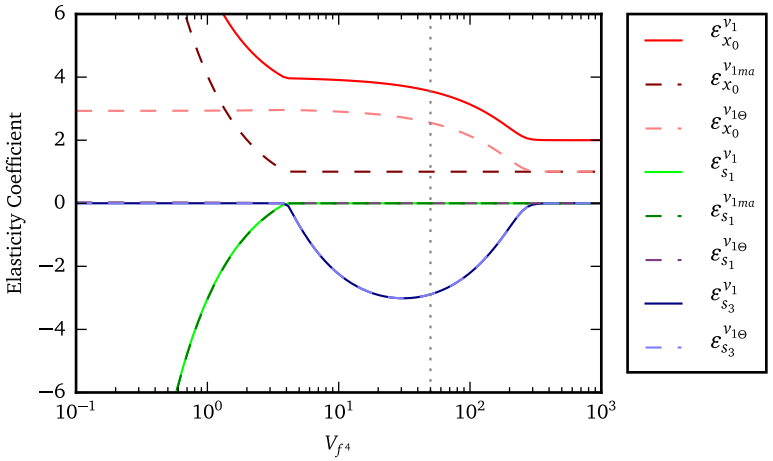


Figure 3.5: An example of a 2D-plot generated by `PySCeSToolbox` and refined using Matplotlib. This publication quality figure was produced by modifying a `PySCeSToolbox` figure using standard Python and Matplotlib functionality. The elasticity coefficients of reaction 1 in the example pathway with respect to its substrate X_0, product S_1 and inhibitor S_3 are shown as functions of V_{f4}, represented by solid lines. Additionally, the elasticity coefficients of the binding (Θ) and mass-action (v_{ma}) components that constitute the total elasticity coefficient are also included as dashed lines. The dotted vertical line indicates the reference value of V_{f4}.

3.7 Discussion

In this chapter we have presented a software toolbox that implements previously published metabolic analysis frameworks for use in model analysis. The tools provided are `RateChar`, `SymCa`, and `ThermoKin`, which are respectively used to perform GSDA [15], symbolic control analysis [23], and investigations into thermodynamic and kinetic aspects of enzyme-catalysed reactions [16]. Until now these frameworks have, at best, been available as prototype software implementations with relatively minimal capabilities. In contrast, `PySCeSToolbox` has been developed in order to provide a robust and feature-rich model analysis platform.

One of the main strengths of `PySCeSToolbox` stems from its design as a library and its integration with the Python ecosystem. Over the last two decades the Python programming language has steadily become a major player within the scientific computing landscape [157, 181, 182], competing with both commercial packages, such as `Matlab` [89] and `Mathematica` [90], and other open-source alternatives, such R [183] and `GNU Octave` [184]. This popularity has, to a large extent, been driven by the availability of numerous excellent libraries for performing scientific analysis [182]. While `PySCeSToolbox` has been developed using a number of these scientific libraries, being designed as library itself allows for incorporating third-party libraries in the modelling experiments. In this sense, while

PySCeSToolbox can be used in a stand-alone manner with PySCeS, it is meant to be able to seamlessly leverage any external functionality. Additionally, the library-style design promotes the development of modelling scripts, thus removing the need for user guided experimentation. Aside from the convenience provided by this practice, it also facilitates reproducibility and repeatability. While Python is, in many cases, slower in terms of execution speed compared to compiled languages, it makes up for this by facilitating a faster development time through simplified syntax and access to external functionality [185]. Furthermore, execution speed can be increased by making use of NumPy array operations [179] rather than equivalent tight loops, or by converting especially slow functions to compiled Cython functions [186].

PySCeSToolbox has also been developed with user-experience and flexibility in mind. The main tools of PySCeSToolbox follow a consistent workflow where functions are similarly named, organised, and accessed; and the generated results of each tool are stored and accessed in a similar manner using the same types of data structures. After generating the main results (i.e. parameter scans, symbolic control coefficient expressions, separated rate equation terms), each of RateChar, SymCa, and ThermoKin provide similar additional functionality such as parameter scans and interactive plotting of results. Finally, multiple tools can be utilised at the same time on a single model, with changes made in model parameters being reflected in the results provided by each tool simultaneously. These features provide not only a consistent user experience, but, as we will discuss below, a means for utilising these tools in an integrated manner for metabolic model analysis.

Supply-demand analysis is arguably the most well known of the frameworks presented here, with the concepts developed by its authors having become increasingly well accepted within systems biology. Specifically, the idea that flux is often controlled by the demand for an intermediate, rather than the supply, is encountered in a variety of metabolic systems, as extensively reviewed by Morandini [129]. However, while SDA has been utilised in a number of experimental metabolic studies in the past (e.g. [28, 29, 125, 128]), to our knowledge GSDA has only been applied in a single metabolic modelling study (refer to Chapter 4). We applied GSDA in order to quantify the regulatory importance of the routes of interaction between intermediates and their reaction blocks in two metabolic models [31]. One of our main findings was that, in a model of aspartate-derived amino-acid synthesis [85], aspartate-semialdehyde unexpectedly affected the rate of its supply block mainly via interactions with its four different demand blocks rather than via product inhibition of its supply enzyme. We could quantify each of the different routes of regulation in terms of their relative importance in the reference system, as well as for two knockout systems; thus demonstrating that some routes of regulation are more important than others.

The results of the above-mentioned study were generated using functionality provided

by `RateChar` and `PySCeS`. In principle, the same results could have been generated using only `PySCeS` and its MCA and parameter scan functionality; however, this would have been a much more involved and cumbersome process that would have required the development of relatively complicated modelling scripts. By making use of the plainly-presented functionality provided by `RateChar`, we could generate results and figures using short and simple scripts, which could, in turn, easily be repurposed to perform the same experiments on the different knockout models. Moreover, we could easily compile and share the scripts and procedures for generating the results presented in our publication in the form of a `Jupyter Notebook`, included as an appendix, thus contributing to the drive for transparency and open data in science [161].

The concepts of symbolic control analysis and the related control pattern analysis have existed for at least 30 years. In 1986, Hofmeyr [118] presented the idea of determining control coefficient expressions without the need for symbolic matrix inversion. This diagrammatic approach was developed in response to the lack of software with symbolic matrix inversion capabilities at the time, and as an alternative to the error-prone and difficult process of performing such inversions by hand. Relatively soon after, two software tools were published that could indeed produce symbolic control coefficient expressions [22, 120]; however, to our knowledge they have not been utilised in any published work that explores the control structure of metabolic pathways in the same way suggested here. Examples of simple derivations of symbolic control coefficient expressions do, however, appear in the literature, but their use is limited to explaining general principles of control in metabolic system (e.g. [14, 187–189]), or to relatively small models in which these expressions were used to calculate numeric control coefficient values for different elasticity coefficient values (e.g. [27, 122]). The latest version of `SymCa` presented in this chapter, however, is a significant improvement over all previous attempts to implement symbolic control analysis. The software tools published in the early 1990s were severely limited in terms of their functionality. One of these tools, for instance, required the user to manually select a reference flux [22], whereas the other could only be applied to linear pathways with a finite number of branches leading from the main pathway [120]. They were also rather cumbersome to use compared to `SymCa`. These tools were therefore not suitable for symbolic control analysis of metabolic models of any significant scale (a fact that was compounded by the relatively limited computing power available at the time). Aside from the previously mentioned benefits of the design and platform of `PySCeSToolbox`, the new `SymCa` also offers a greatly improved compatibility and speed advantage over the previous version. This older version relied completely on `Maxima` to perform all symbolic computation and expression manipulation [23, 24]. This was a much more unwieldy and unreliable approach than is followed here, as it required the development of a

dedicated interface between SymCa and Maxima that magnified the possibility for introducing errors in the software by increasing the size and complexity of the code.

The original incarnation of SymCa does, however, represented the first substantial development of symbolic control analysis since the work of the 1990s, as well as its first application to the study a metabolic models. In his thesis [24], Akhurst investigated a number of metabolic systems using this framework. Notably, through a symbolic control analysis of sucrose accumulation in sugar cane (*Saccharum officinarum*) culm, he provided quantitative evidence for the decrease in futile cycling in maturing internodes originally observed by Uys *et al.* [123]. These results would not have been achievable using any previously available software without significant effort. Recently, the latest version of SymCa has also been used to explain the necessary conditions for ultrasensitivity in redoxin networks of *Escherichia coli* by processing generated control coefficient expressions using functionality provided by SymPy [190].

In 2009, Yuan *et al.* [30] developed a detailed kinetic model of central nitrogen metabolism in *Escherichia coli*. They used the techniques of Hofmeyr [25] (which would later form the basis of the framework of Rohwer and Hofmeyr [16]) to investigate the effect of a nitrogen shift on the elasticity coefficients of the GOGAT and aminotransferase enzymes. Before nitrogen upshift they found that the elasticity coefficient of GOGAT with respect to glutamine was 0.37, which decreased after nitrogen upshift. They also investigated the effect of the nitrogen shift on the elasticity coefficients of the individual kinetic and thermodynamic terms (called the mass-action term here as per [25]). Their results indicated that GOGAT was saturated with respect to glutamine and that the flux through this reaction was insensitive to the concentration of this substrate. In this study, in-house developed code to was used to perform all simulations and modelling experiments. This means that code had to be developed specifically for this single model in order to separate the rate equations into their respective kinetic and thermodynamic terms, to subsequently symbolically partially differentiate these terms with respect to substrates and products to obtain elasticity coefficient expressions, and to substitute the numeric concentration and parameter values into the elasticity coefficient expressions to obtain their values for the different conditions. Clearly, the use of ThermoKin would have saved these authors a significant amount of work, had this software been available at the time.

Other authors have proposed different ways to separate the terms of a rate equation to reveal the contributions of various effects on the rate and control of reactions. Both Noor *et al.* [180] and Morandini [129], for example, suggest moving the K_m constant of the rate capacity term to the binding term. Liebermeister *et al.* [191], on the other hand, include the forward and reverse k_{cat} terms of their suggested modular rate laws for enzymatic reactions with the

turnover number term T_r (which is equivalent to the mass-action term of [16]). While these approaches differ from those suggested by Rohwer and Hofmeyr [16], `ThermoKin` accepts rate equations separated in any arbitrary manner, as long as these term separations are valid (i.e. multiplication of these terms yields the original rate equation). Thus `ThermoKin` has the potential to be utilised in ways beyond those suggested in this chapter.

As already mentioned, none of the frameworks implemented in `PySCeSToolbox` have been used in a complementary manner to investigate metabolic systems in any published works. In Chapter 5, however, we extended upon our previous GSDA study [31] (Chapter 4) to include analyses utilising both `SymCa` and `ThermoKin`. There we report on the effects of a perturbation of the redox balance on the control patterns of the pyruvate branch model [192] investigated in [31] (Chapter 4), as well as the effect of this perturbation on the thermodynamic and kinetic aspects of the enzymes of this pathway. Furthermore, we demonstrate how, in some cases, one of these tools can be used to explain the results obtained from another one.

3.8 Conclusion

The software presented in this chapter represents a marriage between modern computational technologies and theoretical analysis frameworks, therefore providing researchers with a new set of tools for investigating metabolic systems. It includes software implementations of three previously conceived, but as of yet arguably neglected, metabolic analysis frameworks. At its core, `PySCeSToolbox` has been developed specifically with the goals of systems biology in mind; that is to say that it is meant to be utilised in the analysis of models of metabolic systems in order to understand how their higher-level system behaviour arises from their molecular components.

To facilitate this goal, we have developed this software to provide a streamlined and accessible user experience, as well as a flexible and powerful approach to performing modelling experiments. This is achieved in four ways. The first, and most important, is the rich feature set provided by `PySCeSToolbox` that allows users to perform a variety of analyses and to explore the properties of a metabolic model. Second is the use of a unified and simple design; each component functions in very much the same way from the user's perspective and is thus easy to learn and to apply. Thirdly, it combines a powerful scripting paradigm and a simple-to-use graphical user interface, thus making it an accessible and viable tool for users over a spectrum of technical skill levels. Finally, it fits into the scientific Python ecosystem, thus providing users with additional flexibility through access to third-party tools.

In the future we envision the use of `PySCeSToolbox` in the analysis of previously pub-

lished metabolic models in order to reach a better understanding of the systems they represent. Specifically, we intend to utilise this software as part of an ongoing study regarding the properties of redoxin networks (see e.g. [193, 194]). Other work will include the further development of the software itself. Currently `SymCa` utilises `Maxima` to perform symbolic expression factorisation due to its superiority in this regard compared to `SymPy`. However, as the underlying technology of `SymPy` improves, we hope to remove the `Maxima` dependency (see also Section 6.3). In the future the pathway scheme functionality will also be improved by adding automated network layout capabilities. Finally, ongoing support will be provided by addressing bug reports or feature requests through `https://github.com/PySCeS/PyscesToolbox`.

Chapter 4

Tracing Regulatory Routes in Metabolism Using Generalised Supply-Demand Analysis[1]

4.1 Summary

Background

Generalised supply-demand analysis is a conceptual framework that views metabolism as a molecular economy. Metabolic pathways are partitioned into so-called supply and demand blocks that produce and consume a particular intermediate metabolite. By studying the response of these reaction blocks to perturbations in the concentration of the linking metabolite, different regulatory routes of interaction between the metabolite and its supply and demand blocks can be identified and their contribution quantified. These responses are mediated not only through direct substrate/product interactions, but also through allosteric effects. Here we subject previously published kinetic models of pyruvate metabolism in *Lactococcus lactis* and aspartate-derived amino acid synthesis in *Arabidopsis thaliana* to generalised supply-demand analysis.

Results

Multiple routes of regulation are brought about by different mechanisms in each model, leading to behavioural and regulatory patterns that are generally difficult to predict from

[1] This chapter has been published: Christensen, Hofmeyr & Rohwer, *BMC Syst. Biol.*, 2015, 9:89 [31] and is reproduced here.

simple inspection of the reaction networks depicting the models. In the pyruvate model the moiety-conserved cycles of ATP/ADP and NADH/NAD^+ allow otherwise independent metabolic branches to communicate. This causes the flux of one ATP-producing reaction block to increase in response to an increasing ATP/ADP ratio, while an NADH-consuming block flux decreases in response to an increasing NADH/NAD^+ ratio for certain ratio value ranges.

In the aspartate model, aspartate semialdehyde can inhibit its supply block directly or by increasing the concentration of two amino acids (Lys and Thr) that occur as intermediates in demand blocks and act as allosteric inhibitors of isoenzymes in the supply block. These different routes of interaction from aspartate semialdehyde are each seen to contribute differently to the regulation of the aspartate semialdehyde supply block.

Conclusions

Indirect routes of regulation between a metabolic intermediate and a reaction block that either produces or consumes this intermediate can play a much larger regulatory role than routes mediated through direct interactions. These indirect routes of regulation can also result in counter-intuitive metabolic behaviour. Performing generalised supply-demand analysis on two previously published models demonstrated the utility of this method as an entry point in the analysis of metabolic behaviour and the potential for obtaining novel results from previously analysed models by using new approaches.

4.2 Background

The primary concern of molecular biology is to identify and characterise the individual components of biological systems. By focussing on the component level, this approach has generated an enormous amount of knowledge, but at the expense of disregarding emergent phenomena that arise from the multitude of interactions between these components. One way of overcoming this limitation is to construct, and subsequently study, mathematical models of these biological systems. This technique has become increasingly common, with models describing systems ranging in complexity from metabolic subsystems to genome-scale reconstructions of metabolism [195, 196] being available on various online databases [159, 160, 197]. More recently the goal of building a silicon cell [198] was arguably realised with the construction of a whole-organism model of *Mycoplasma genitalium* [18]. With models growing in size and complexity, approaching that of the modelled systems themselves, it has become more difficult to extract biological knowledge and understanding from them without extensive analysis.

Model construction is therefore only the first step in the study of biological systems using the 'modelling approach'.

Generalised supply-demand analysis (GSDA) is a conceptual framework that views metabolic pathways as a molecular economy [15]. It is built on the principles of metabolic control analysis (MCA) [9, 199], which is itself a framework that allows for the quantification of the control that any step in the system exercises over the variable steady-state properties such as fluxes or intermediate metabolite concentrations. The basic procedure of supply-demand analysis is to divide a metabolic pathway into separate reaction blocks around a chosen variable metabolite by fixing its concentration; the 'generalised' in GSDA implies that this is done in turn for each metabolite in the system. The supply and demand blocks, which respectively produce and consume the fixed metabolite, are subsequently treated as separate metabolic units, and MCA is performed on each reaction block. This approach allows for the identification of certain regulatory features of the pathway and for the quantification of the behaviour of the reaction blocks towards to changes in the concentration of the fixed metabolite. One such feature that GSDA helps to identify and quantify is the effect of different routes of interaction between the variable metabolites and their supply and demand reaction blocks.

The simplest way that reaction blocks can interact is through the common intermediate that links them, which can serve as either substrate or product or allosteric effector of supply and demand enzymes. If the only interactions are via the linking metabolite the situation is easy to analyse. However, it is possible that reaction blocks also interact indirectly through allosteric effects of a metabolite in one reaction block on an enzyme in the other reaction block; such a situation is quite common and it is here that GSDA is particularly useful in that it dissects all the routes of communication between supply and demand, both direct and indirect.

Another common situation is where cofactor cycles such as ATP–ADP or NADH–NAD^+ link supply and demand reaction blocks. These cycles are called moiety-conserved cycles because they interconvert different forms of a chemical subgroup or moiety, the total concentration of which remains constant. For example, in the ATP/ADP cycle the moiety is ADP, which is interconverted between its free and its phosphorylated form. When there is no *de novo* synthesis or degradation of the ADP-moiety during the time-scale of observation, the sum of ADP and ATP remains constant even while their individual concentrations change, and the cycle is therefore moiety-conserving. In supply-demand analyses of such cycles the relevant variable that links the supply and demand reaction blocks is not a single concentration but rather a concentration ratio such as ATP/ADP or NADH/NAD^+. These cycles usually form metabolic hubs where many functionally distinct pathways are integrated; an analysis of the interplay between supply and demand around these cycles is crucial for our understanding

of metabolic regulation.

In this paper we use GSDA to investigate the regulatory effects brought about by multiple routes of interaction between supply and demand reaction blocks. We have chosen to analyse models of two metabolic pathways that differ from each other in terms of their regulatory mechanisms. The first is a model of pyruvate metabolism in lactic acid bacteria [86] where multiple interactions are brought about through the moiety-conserved cycles of ATP/ADP and NADH/NAD^+. We show that changes in the ratios of the different forms of a moiety can cause counter-intuitive responses in reaction block fluxes. The second is a model of aspartate-derived amino acid synthesis in *Arabidopsis thaliana* [85] in which we analyse the routes of interaction brought about by allosteric effectors in combination with multiple isoenzymes. Here we explore the functions of the various isoenzymes and construct a map that shows the importance of the routes of regulation between a fixed metabolite and its supply block. In both models the importance of each route of interaction originating from a change in the fixed metabolite will be quantified and compared, illustrating that the direct route of interaction is not necessarily the most important. More generally we demonstrate the utility of investigating previously published models with a new analytic tools such as GSDA.

4.3 Methods

4.3.1 Metabolic control analysis

Metabolic control analysis (MCA) is a form of sensitivity analysis in which the control properties of a steady-state metabolic system are quantified in terms of the responses of the system fluxes and metabolite concentrations to perturbations in the rates of the reactions [9, 199]. Because this framework plays a central role in generalised supply-demand analysis [15], we here define its three main coefficients and their relation to each other.

An *elasticity coefficient* describes the sensitivity of a reaction rate towards a change in any entity x that can affect that rate directly, such as a substrate, product, modifier or enzyme parameter. It is therefore a property local to a particular reaction and is defined as the ratio of the relative change in the rate of reaction i, v_i, to the relative change in x:

$$\varepsilon_x^{v_i} = \frac{\partial \ln v_i}{\partial \ln x} \tag{4.1}$$

A *control coefficient* describes the sensitivity of a steady-state system variable, such as flux or concentration, towards a change in a local reaction rate. This coefficient is a systemic property that depends not only on the properties of the perturbed reaction but also on those of the other reactions and the topology of the network structure of the entire pathway. The

flux-control coefficient is defined as the ratio of the relative change in a flux, J, to the relative change in a reaction rate, v_i:

$$C^J_{v_i} = \frac{d \ln J}{d \ln v_i} \tag{4.2}$$

Concentration-control coefficients are defined similarly, the flux being replaced by a metabolite concentration. The use of a total derivative signifies that the entire system is allowed to relax to a new steady state after the perturbation in v_i.

A *response coefficient* differs from a control coefficient in that it is defined with respect to a change not in a local reaction rate but in a system parameter, such as enzyme concentration or the fixed concentration of metabolite external to the system. A flux-response coefficient is defined similarly to a control coefficient as the ratio of the relative change in a flux, J, to the relative change in parameter x:

$$R^J_x = \frac{d \ln J}{d \ln x} \tag{4.3}$$

Again, in a concentration-response coefficient metabolite concentration replaces flux.

The so-called partitioned response (or combined-response) equation describes the relationship between these three coefficients:

$$R^J_x = C^J_{v_i} \varepsilon^{v_i}_x \tag{4.4}$$

The overall flux-response to a perturbation in parameter x is channelled through the reaction i directly affected by x: the effect of δx on v_i is described by $\varepsilon^{v_i}_x$, and the resulting change δv_i then propagates through the system resulting in a change in flux described by $C^J_{v_i}$. If the parameter x affects more than one reaction, the overall flux-response is given by

$$R^J_x = \sum_i C^J_{v_i} \varepsilon^{v_i}_x \tag{4.5}$$

for all reactions i that are influenced directly by x.

4.3.2 Generalised supply-demand analysis

Generalised supply-demand analysis (GSDA) is an extension of metabolic supply-demand analysis [14]. In GSDA a metabolic pathway is partitioned into reaction blocks around each variable metabolite, as shown for the metabolite P in the linear pathway in Figure 4.1A. The producing and consuming blocks of this intermediate are termed the supply and demand blocks, respectively. The concentration of each variable is fixed and, in turn, varied over a range to generate combined rate characteristic plots [25] of the supply and demand blocks linked to the metabolite as shown in Figure 4.1B. The response coefficients of the supply and demand blocks towards the linking metabolite are calculated, along with the elasticity

coefficients of the reactions in these blocks that are directly connected to the intermediate (i.e., the last reaction in the supply block and the first in the demand block).

This approach can be used to determine which block has the most control over the system flux and which block determines the degree of homoeostasis of the linking intermediate by comparing the response coefficients (or gradients of the rate characteristics at steady-state) of the supply and demand blocks towards the intermediate. In the case where

$$\left| \frac{R_p^{J_{\text{supply}}}}{R_p^{J_{\text{demand}}}} \right| > 1 \tag{4.6}$$

as in Figure 4.1B, the flux is predominantly controlled by the demand block, while in the opposite case where the ratio is < 1, the flux is controlled by the supply. The degree of homoeostatic maintenance of the P concentration (denoted by p) depends on the value of $R_p^{J_{\text{demand}}} - R_p^{J_{\text{supply}}}$; the larger this value the smaller the absolute value of the concentration-control coefficients of the supply and demand blocks on p, and the better its homoeostasis [14, 200].

4.3.3 Multiple routes of interaction

GSDA can also be used to identify the different routes of interaction between an intermediate and a reaction block and to quantify the relative contribution of each of these routes to the total response. In Figure 4.1C an additional interaction of P with its supply block occurs through the allosteric inhibition of reaction 1 by P. The flux-response coefficient of the supply block to P is now the sum of two terms, called partial response coefficients:

$$\begin{aligned} R_p^{J_{\text{supply}}} &= {}^{\nu_1}R_p^{J_{\text{supply}}} + {}^{\nu_2}R_p^{J_{\text{supply}}} \\ &= C_{\nu_2}^{J_{\text{supply}}} \varepsilon_p^{\nu_2} + C_{\nu_1}^{J_{\text{supply}}} \varepsilon_p^{\nu_1} \end{aligned} \tag{4.7}$$

According to the partitioned-response property (Equation 4.5), each partial response coefficient is the product of an elasticity coefficient and a control coefficient. These control coefficients represent the sensitivities of flux local to the supply block, and must be distinguished from the flux-control coefficients of the full supply-demand system, i.e., $C_{\nu_i}^{J_{\text{supply}}} \neq C_{\nu_i}^{J}$. The partial responses can be represented visually in the form of a rate characteristic plot as shown in Figure 4.1D.

Whereas in Figures 4.1A and 4.1C the metabolite P is the sole link between the supply and demand blocks, it is of course also possible, as in Figure 4.1E, for an intermediate within one reaction block to provide a link to the other reaction block. Here, reaction 1 is inhibited by S_2, so that a change in P will be transmitted via reaction 3 in the demand block to S_2,

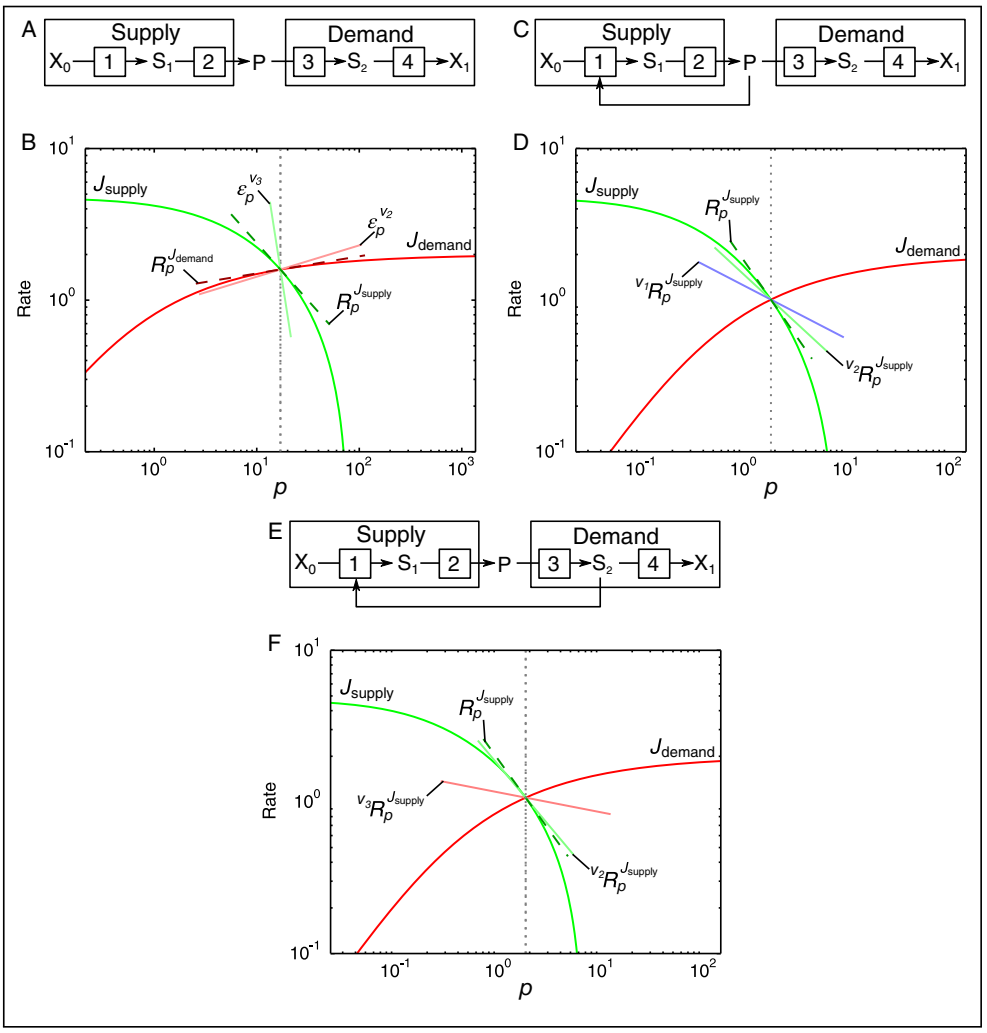


Figure 4.1: An example of generalised supply-demand analysis of three metabolic systems. **(A)** An example of a simple linear pathway partitioned into supply and demand blocks around intermediate P. **(B)** A rate characteristic plot that shows how the fluxes local to the supply and demand blocks of P respond to a change in p over a large concentration range. The vertical dotted line indicates the steady-state concentration of P ($\bar{p}$), while the steady-state flux of the system as a whole obtains where the rates of the supply and demand blocks intersect. Elasticity coefficients of the reactions in the supply and demand blocks that interact directly with P (reactions 2 and 3 respectively) are indicated with solid lines while response coefficients of the blocks towards p are indicated by dashed lines. **(C)** The pathway in **(A)** with the addition of allosteric inhibition of enzyme 1 by P, which creates an additional direct route of interaction between P and its supply block via reaction 1. **(D)** The rate characteristic plot for the supply and demand blocks of the intermediate P in **(C)**. **(E)** The pathway in **(A)** with the addition of allosteric inhibition of enzyme 1 by S_2, which creates an indirect route of interaction between P and its supply block via reaction 3. **(F)** The rate characteristic plot for the supply and demand blocks of the intermediate P in **(E)**. In **(D)** and **(F)** only the total and partial response coefficients of the supply block towards p are shown and the slopes of the partial response coefficients (solid lines) add up to that of the total response coefficient (dashed line).

affecting its concentration, which, in turn, affects reaction 1. The overall effect (as shown in Figure 4.1F) is similar to that in Figures 4.1B and 4.1D, except that P also affects the supply block indirectly via S_2. This indirect effect of P via S_2 is made explicit in the second right-hand term of the following expression:

$$\begin{aligned} R_p^{J_{\text{supply}}} &= {}^{v_1}R_p^{J_{\text{supply}}} + {}^{v_3}R_p^{J_{\text{supply}}} \\ &= C_{v_2}^{J_{\text{supply}}}\varepsilon_p^{v_2} + C_{v_3}^{J_{\text{supply}}}\varepsilon_p^{v_3} \\ &= C_{v_2}^{J_{\text{supply}}}\varepsilon_p^{v_2} + C_{v_1}^{J_{\text{supply}}}\varepsilon_{S_2}^{v_1}C_{v_3}^{S_2}\varepsilon_p^{v_3} \end{aligned} \tag{4.8}$$

The fixed metabolite can also interact with a particular reaction block through indirect stoichiometric linkages. In this case a change in the fixed metabolite concentration is transmitted via one reaction block to another through stoichiometric connections in the rest of the network, in a similar manner to the previously described allosteric interaction (Figure 4.1E). The difference here is that, instead of an allosteric interaction, metabolites and reactions can link both reaction blocks via a stoichiometric route that does not involve the fixed metabolite. The members of a moiety-conserved cycle (discussed below) are an example of intermediates that can link reaction blocks in this way because of their stoichiometric involvement in numerous reactions at various points in the network.

4.3.4 Moiety-conserved cycles

Moiety-conserved cycles require special consideration in GSDA as the total concentration of the members of the conserved cycle must remain constant. The individual member concentrations are therefore not free to vary in the same way as other metabolites.

In order to perform parameter scans on the members of the moiety-conserved cycles without breaking moiety conservation, the individual members of a cycle can be substituted with a single metabolite representing their ratio to one another. The concentrations of the members of each cycle are calculated using the total moiety concentration and the value of the ratio. Using the ATP/ADP moiety-conserved cycle as an example, with ϕ_A representing the ratio of ATP to ADP and C_A the total moiety concentration, the equations below illustrate this principle:

$$\begin{aligned} \phi_A &= \frac{[\text{ATP}]}{[\text{ADP}]} \\ C_A &= [\text{ATP}] + [\text{ADP}] \\ \therefore [\text{ATP}] &= \frac{\phi_A \cdot C_A}{1+\phi_A} \quad \text{and} \quad [\text{ADP}] = \frac{C_A}{1+\phi_A} \end{aligned}$$

4.3.5 Software

The Python simulator for cellular systems (PySCeS) [91] together with the RateChar [15] module that forms part of the PySCeSToolbox package (https://github.com/PySCeS/PyscesToolbox) was used to perform the modelling experiments and metabolic control analysis and to generate the resulting rate characteristic plots. RateChar automatically performs supply-demand analysis and produces rate-characteristic plots for each metabolite in a metabolic model.

4.3.6 Models

4.3.6.1 Pyruvate branch metabolism

To investigate the effects of multiple routes of regulation through moiety-conserved cycles we used a kinetic model of pyruvate metabolism in lactic acid bacteria. The model was originally constructed by Hoefnagel *et al.* [86] and retrieved from JWS online [160] in the PySCeS model descriptor language format [91, 201]. The structure of the model is outlined in Figure 4.2. This model was chosen for our investigation on the basis of its three different moiety-conserved cycles that interact with a variety of reactions across different branches in the pathway.

The members of the ATP/ADP, acetyl-CoA/CoA and NADH/NAD^+ conserved moieties were treated as outlined previously, with the symbols ϕ_A, ϕ_C and ϕ_N representing the metabolite ratios. The values of ϕ_A and ϕ_N were fixed and varied over the ranges shown in Table 4.1.

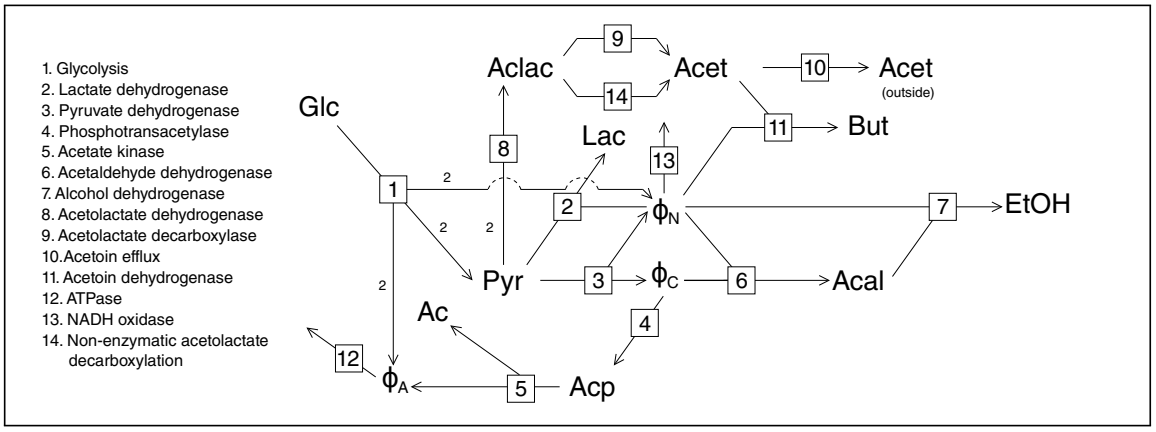


Figure 4.2: The pyruvate branch pathway as defined by Hoefnagel *et al.* [86]. Reactions are numbered according to the key. The stoichiometry of each reaction is 1 to 1, except for reaction 1 where $Glc + 2ADP + 2NAD^+ \rightarrow 2Pyr + 2ATP + 2NADH$ and reaction 8 where $2Pyr \rightleftharpoons Aclac$. Intermediates are abbreviated as follows: Ac: acetate; Acal: acetaldehyde; Acet: acetoin; Aclac: acetolactate; Acp: acetyl phosphate; Glc: glucose; Lac: lactate; But: 2,3-butanediol; Pyr: pyruvate; EtOH: ethanol.

Table 4.1: Pyruvate metabolism model scan ranges. The ranges over which the variable metabolite concentration ratios of the pyruvate metabolism model [86] were varied to generate Figures 4.4–4.8.

Metabolite	Scan range	Steady-state ratio
ϕ_A	0.06–457.51	5.08
ϕ_N	0.0002–1.77	0.02

Together with the results from metabolic control analysis these data were used to generate Figures 4.4–4.8.

4.3.6.2 Aspartate metabolism

A kinetic model of aspartate-derived amino acid synthesis in *Arabidopsis* was used to investigate the effects of multiple routes of regulation brought about by allosteric effectors and multiple isoenzymes. The model was originally constructed by Curien *et al.* [85] in Berkeley Madonna format and was translated to the PySCeS MDL format. The structure of this pathway is outlined in Figure 4.3.

For this case study the focus was to identify and quantify the different routes of regulation of aspartate-semialdehyde (ASA) with its supply block for the wild type as well as for knockouts of AKI and AKII. ASA, lysine (Lys) and threonine (Thr) were fixed and varied over the ranges shown in Table 4.2 and used, together with results from MCA, to generate Figures 4.9–4.13.

The knockouts of AKI, AKII and both AKI and AKII were simulated by setting their respective enzyme concentrations to zero in the model.

4.3.6.3 Model code

Code for both models is provided in PySCeS model descriptor language (Additional files 2 and 3) and in the standard Systems Biology Markup Language (SBML) format [202] (Additional

Table 4.2: Aspartate metabolism model scan ranges. The ranges over which the variable metabolite concentrations of the aspartate-derived amino acid synthesis pathway model [85] were varied to generate Figures 4.9 - 4.12.

Metabolite	Scan range (μM)	Steady-state conc. (μM)
ASA	0.024–32.97	0.96
Thr	19.80–4453.93	296.93
Lys	4.61–1037.38	69.16

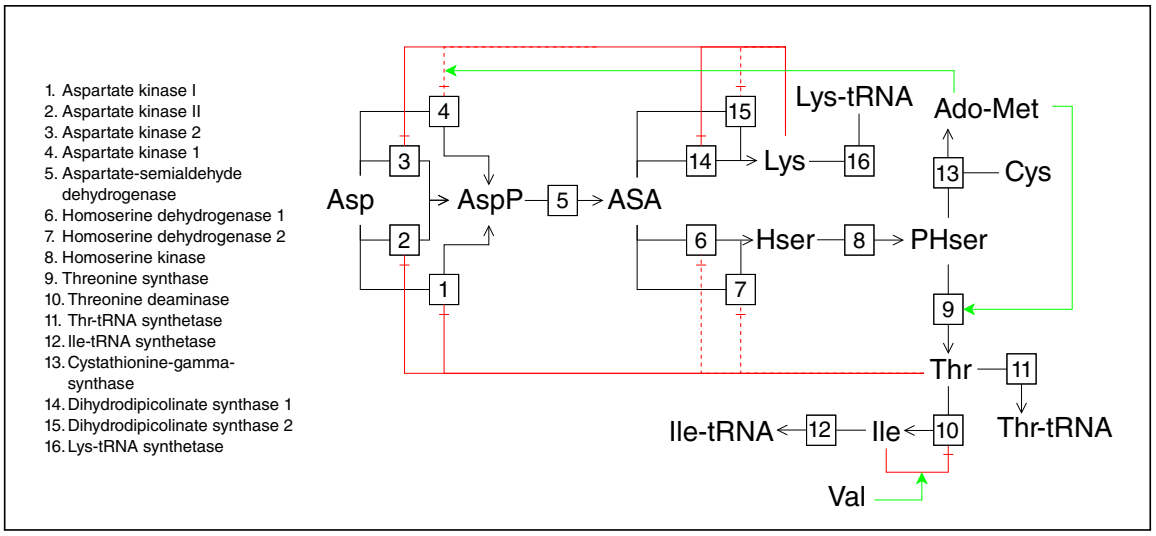


Figure 4.3: The aspartate-derived amino acid synthesis pathway as defined by Curien *et al.* [85]. Reactions are numbered according to the key. Green lines ending with solid arrows indicate activation of a reaction or potentiation of an allosteric effect, while red lines ending with daggers indicate inhibition of reactions or damping of an allosteric effect. Strong allosteric effects are indicated with solid lines, while weak effects are shown with dashed lines. The stoichiometry of each reaction is 1 to 1. Intermediates are abbreviated as follows: Ado-Met: S-adenosylmethionine; ASA: aspartate-semialdehyde; Asp: aspartate; AspP: aspartyl phosphate; Cys: cysteine; Hser: homoserine; Ile: isoleucine; Lys: lysine; PHser: phosphohomoserine; Thr: threonine; Val: valine

files 4 and 5). An IPython notebook containing instructions and scripts to reproduce the analyses from this paper is provided as Additional file 6.

4.4 Results

4.4.1 Regulatory connections via moiety-conserved cycles

The pyruvate branch model in Figure 4.2 contains three moiety conserved cycles, ATP/ADP, $NADH/NAD^+$ and acetyl-CoA/CoA, with their members modelled as variable species. These species take part in a number of reactions across the three main branches leading from pyruvate, thereby enabling the branches to communicate with each other. This model is therefore an ideal candidate for investigating the type of behaviour that can occur due to the presence of moiety-conserved cycles in general, and more specifically due to the ATP/ADP and $NADH/NAD^+$ cycles, which are ubiquitous in metabolism. In this section we show how the presence of ATP/ADP and $NADH/NAD^+$ causes unexpected and non-monotonic flux response behaviour.

4.4.1.1 Regulatory routes of ATP/ADP

The first conserved moiety we shall investigate is ADP, due to its relatively small number of interactions in this pathway. The 'metabolite' ϕ_A represents the ratio of ATP to ADP. An increase in ϕ_A implies an increase in ATP concentration and a concomitant decrease in ADP concentration within the constraint of their sum being constant. In this pathway ϕ_A is produced by acetate kinase (reaction 5) and a lumped glycolysis pathway (reaction 1), and is consumed by ATPase (reaction 12). The supply and demand blocks for ϕ_A are named according to the numbering of these consuming and producing reactions, i.e., block 1 and block 5 are ϕ_A supply blocks, while block 12 is a ϕ_A demand block, with the fluxes of these blocks symbolised by J_1, J_5 and J_{12} respectively (this naming convention for reaction blocks and their corresponding fluxes is used throughout this paper).

The rate characteristic plot (Figure 4.4A) shows the effect of a change in ϕ_A on its supply and demand blocks. In steady-state all the reaction blocks respond as expected towards increasing ϕ_A, with decreases in J_1 and J_5 and an increase in J_{12}. At ϕ_A values below 0.4, however, there was an *increase* in J_5 in response to increasing product (ATP) and decreasing substrate (ADP) concentration. Considering that acetate kinase is not product-activated, this positive flux response was unexpected.

The source of the flux response of block 5 towards ϕ_A was investigated using partial re-

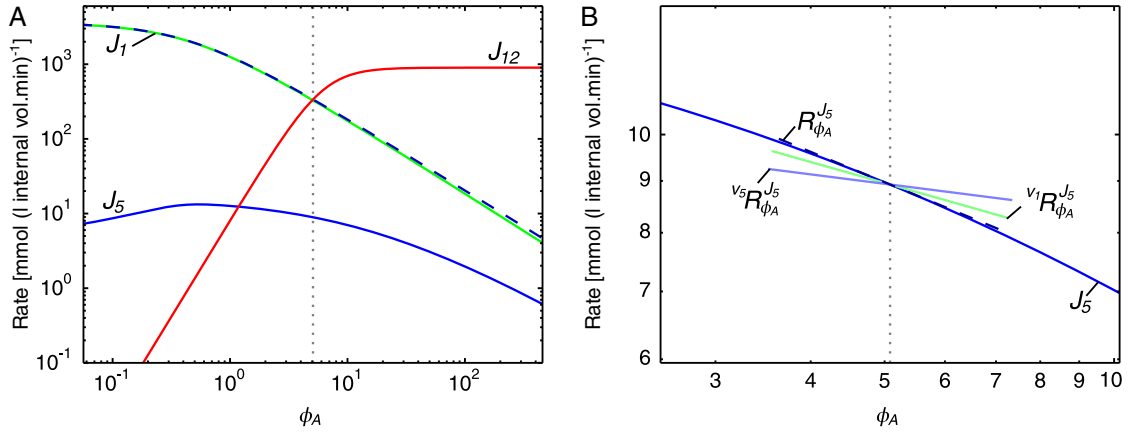


Figure 4.4: Rate characteristic plots of the reaction blocks of ϕ_A in the pyruvate branch model.**(A)** The fluxes of the demand block 12, and the supply blocks 1 and 5 of ϕ_A. The unlabelled dashed curve represents the total supply. **(B)** The rate characteristic plot of the ϕ_A-supply block 5 with partial and total response coefficients indicated as lines intersecting the J_5-curve at the steady-state value of ϕ_A. Partial response coefficients (solid lines) indicate the relative contribution of each route of interaction of ϕ_A with block 5 towards the total response coefficient (dashed line). The steady-state value of ϕ_A is indicated as a vertical dotted line in both **(A)** and **(B)** (see Table 4.1).

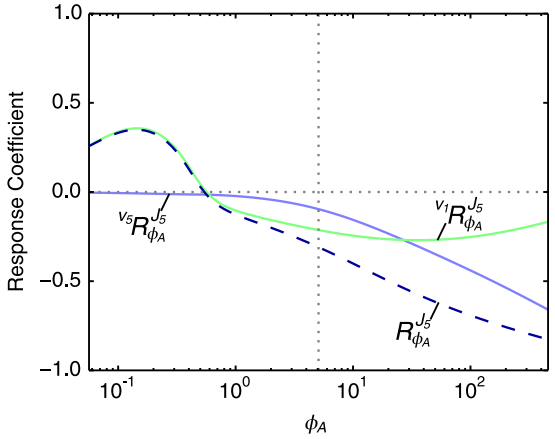


Figure 4.5: Partial and total response coefficients of J_5 towards ϕ_A as a function of ϕ_A. Partial response coefficients (solid lines) indicate the relative contribution of each route of interaction of ϕ_A with reaction block 5 towards the total response coefficient (dashed line) over the ϕ_A-range as indicated in Table 4.1. The steady-state value of ϕ_A is indicated as a vertical dotted line.

sponse coefficients (Figure 4.4B) where a rate characteristic plot revealed that two different routes of interaction are responsible for the behaviour of this reaction block. At the steady-state, the partial response coefficients representing the relative importance of these two routes are both negative, leading to the observed negative total response coefficient of J_5. The first partial response ($^{\nu_5}R^{J_5}_{\phi_A}$) is due to the direct interaction of ϕ_A with block 5 via reaction 5. The second, and more significant, partial response ($^{\nu_1}R^{J_5}_{\phi_A}$) is due to the interaction of ϕ_A with reaction 1, which also has a negative elasticity towards ϕ_A and forms part of both reaction blocks 1 and 5. The negative effect on J_5 due to the decrease in J_1 is a result of the flux relationships between these reaction blocks where a decrease in pyruvate production by block 1 leads to a decrease in flux of all pyruvate consuming reactions.

When considering the partial response coefficients over the complete range of ϕ_A-values (Figure 4.5), we saw that whereas $^{\nu_5}R^{J_5}_{\phi_A}$ is close to zero for ϕ_A-values below 0.4, $^{\nu_1}R^{J_5}_{\phi_A}$ is positive, thereby being solely responsible for the observed increase in J_5 over this range of ϕ_A-values. This positive response can again be traced to the flux relationships between J_1 and J_3, but in this case J_2 and J_8 also play a role. While J_1 does decrease, the fluxes J_2 and J_8 decrease even more, resulting in an increase in J_3. Additionally, a decrease in J_6 causes flux to be diverted to J_4. Both these effects lead to the observed increase in J_5 for this ϕ_A-range (Figure 4.4A).

These results indicate that the indirect route of interaction of ϕ_A with block 5 plays a large role in the regulation of the flux through this block, and is indeed the most prominent regulatory route for ϕ_A-values below 30.

4.4.1.2 Regulatory routes of NADH/NAD$^+$

While notable, the counter-intuitive response to ϕ_A is brought about by only two partial responses due to ATP/ADP acting as an intermediate in only a few reactions in the pathway. The NAD$^+$ moiety, on the other hand, interacts with more reaction blocks than either of the CoA and ADP moieties: ϕ_N is produced by a lumped glycolysis pathway (reaction 1) and pyruvate dehydrogenase (reaction 3), and consumed by lactate dehydrogenase (reaction 2), acetoin dehydrogenase (reaction 11), acetaldehyde dehydrogenase (reaction 6), alcohol dehydrogenase (reaction 7) and NADH oxidase (reaction 13). While block 6 and block 7 are separate demand blocks for ϕ_N, they are also linked by acetaldehyde and therefore their rates are equal at steady-state. In cases where the observed results for these blocks are identical we refer only to block 6 for the sake of brevity. Due to the numerous interactions of ϕ_N in this system, there is potential for complex flux response behaviour.

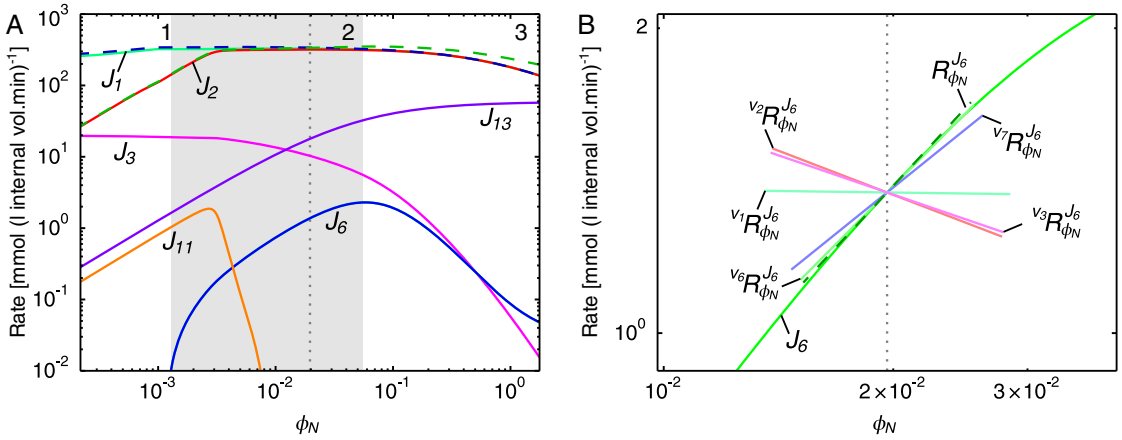


Figure 4.6: Rate characteristic plots of the reaction blocks of ϕ_N in the pyruvate branch model. **(A)** The fluxes of the demand blocks 2, 6, 11 and 13, and the supply blocks 1 and 3 of ϕ_N. The unlabelled dashed curves represent the total supply (blue) and demand (green). **(B)** The rate characteristic plot of the ϕ_N-demand block 6 with partial and total response coefficients indicated as lines intersecting the J_6-curve at the steady-state value of ϕ_N. Partial response coefficients (solid lines) indicate the relative contribution of each respective route of interaction of ϕ_N with the reaction block 6 towards the total response coefficient (dashed line). Note that $^{v_{11}}R^{J_6}_{\phi_N}$ and $^{v_{13}}R^{J_6}_{\phi_N}$ were omitted due to their zero contributions towards $R^{J_6}_{\phi_N}$. The steady-state value of ϕ_N is indicated as a vertical dotted line in both **(A)** and **(B)** (see Table 4.1).

The flux of reaction block 6, J_6 responds non-monotonically to changing ϕ_N (Figure 4.6A), in contrast to the fluxes of the other blocks that ϕ_N reacts with, which respond monotonically as anticipated. In the ϕ_N-range below 0.0012 J_6 is negative (which implies that reaction 6 in Figure 4.2 proceeds in the reverse direction), but becomes less negative as ϕ_N increases;

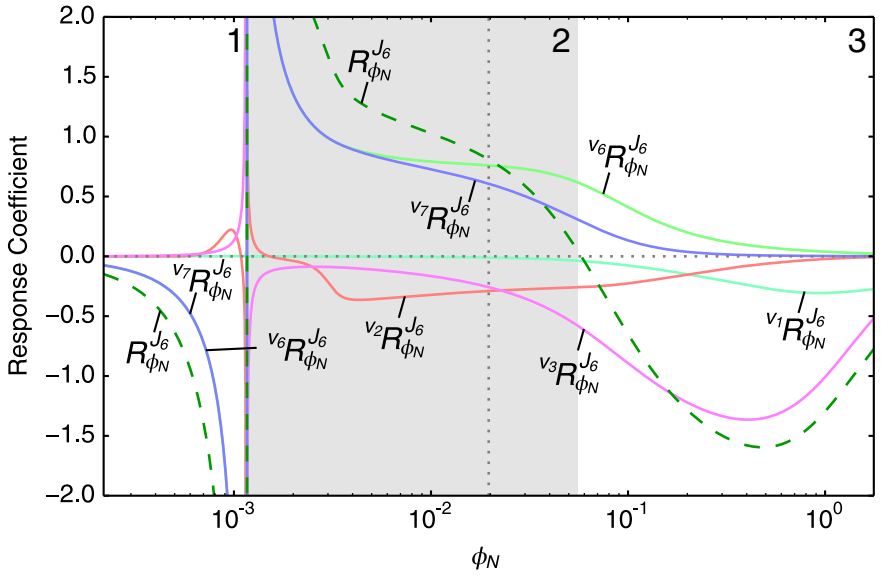


Figure 4.7: Partial and total response coefficients of J_6 towards ϕ_N as a function of ϕ_N. Partial response coefficients (solid lines) indicate the relative contribution of each respective route of interaction of ϕ_N with reaction block 6 towards the total response coefficients (dashed lines) over the ϕ_N-range as indicated in Table 4.1. The steady-state value of ϕ_N is indicated as a vertical dotted line. Note that ${}^{v_{11}}R^{J_6}_{\phi_N}$ and ${}^{v_{13}}R^{J_6}_{\phi_N}$ were omitted due to their zero contributions towards $R^{J_6}_{\phi_N}$.

at $\phi_N = 0.0012$ block 6 is at equilibrium and $J_6 = 0$. In the range $0.0012 < \phi_N < 0.057$ J_6 is positive and increases to a maximum at $\phi_N = 0.057$. In the ϕ_N-range above 0.057 J_6 decreases monotonically. These three ϕ_N-ranges will henceforth be referred to as range 1, 2, and 3 respectively.

As before, partial response coefficients explain the behaviour of reaction block 6. At the steady-state, four routes of interaction of ϕ_N with block 6 contribute significantly to the total response $R^{J_6}_{\phi_N}$, as shown in the rate characteristic plot in Figure 4.6B. The direct interactions via reaction 6 and reaction 7 result in positive partial responses (${}^{v_6}R^{J_6}_{\phi_N}$ and ${}^{v_7}R^{J_6}_{\phi_N}$) due to ϕ_N acting as a substrate for these reactions. On the other hand, the interactions via reactions 2 and 3, represented by ${}^{v_2}R^{J_6}_{\phi_N}$ and ${}^{v_3}R^{J_6}_{\phi_N}$, affect J_6 negatively; by decreasing J_3 via reactions 2 and 3 , ϕ_N decreases ϕ_C, thereby limiting the availability of this additional substrate for reaction 6.

The source of the non-monotonic behaviour of block 6 becomes clear when the partial response coefficients of J_6 towards a range of ϕ_N-values are computed (Figure 4.7). The non-monotonic total response coefficient (the dashed green line) is the sum of multiple partial response coefficients which are themselves non-monotonic, their contributions to the total varying greatly over the ϕ_N-range. There is a singularity at $\phi_N = 0.0012$ between ranges 1

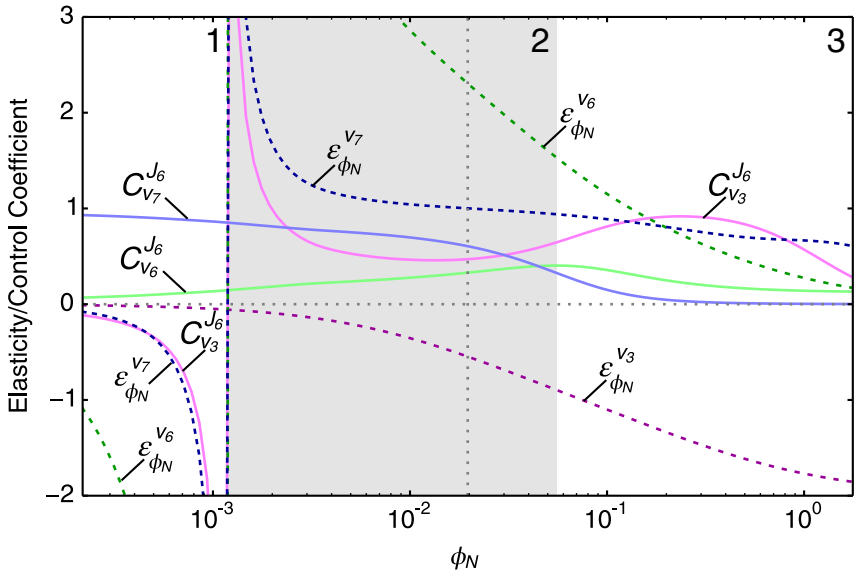


Figure 4.8: The most significant partial response coefficients contributing towards $R^{J_6}_{\phi_N}$ separated into elasticity and control coefficients. Elasticity coefficients (dashed lines) and control coefficients (solid lines) that make up the partial response coefficients of Figure 4.7 are shown. Here $C^{J_6}_{v_6}\varepsilon^{v_6}_{\phi_N} = {}^{v_6}R^{J_6}_{\phi_N}$, $C^{J_6}_{v_7}\varepsilon^{v_7}_{\phi_N} = {}^{v_7}R^{J_6}_{\phi_N}$ and $C^{J_6}_{v_3}\varepsilon^{v_3}_{\phi_N} = {}^{v_3}R^{J_6}_{\phi_N}$. The steady-state value of ϕ_N is indicated as a vertical dotted line (see Table 4.1).

and 2 which correlates with the equilibrium state that block 6 goes through when J_6 changes direction.

In ranges 1 and 2 the total response coefficient behaviour is determined mostly by ${}^{v_6}R^{J_6}_{\phi_N}$ and ${}^{v_7}R^{J_6}_{\phi_N}$ as the values of the other partial response coefficients are low and undergo little change. In range 3, however, a slightly more complex interplay of effects brings about total response behaviour. Here the decline of the total response coefficient and its subsequent reversal of sign was caused by the increase in magnitude of the negative partial response coefficient ${}^{v_3}R^{J_6}_{\phi_N}$ and the decrease in magnitude of ${}^{v_6}R^{J_6}_{\phi_N}$ and ${}^{v_7}R^{J_6}_{\phi_N}$.

By separating the partial response coefficients into their elasticity and control coefficient components according to the partitioned response equation (Equation 4.5), we obtain a clearer view of the role of local versus systemic effects in bringing about the flux response. The control and elasticity coefficients that make up ${}^{v_6}R^{J_6}_{\phi_N}$, ${}^{v_7}R^{J_6}_{\phi_N}$ and ${}^{v_3}R^{J_6}_{\phi_N}$, i.e, the partial responses that make the largest contribution to $R^{J_6}_{\phi_N}$, are shown in Figure 4.8 (see legend for their partitioned response equations). Since elasticity coefficients tend to infinity at equilibrium, $\varepsilon^{v_6}_{\phi_N}$ and $\varepsilon^{v_7}_{\phi_N}$ largely determine ${}^{v_6}R^{J_6}_{\phi_N}$ and ${}^{v_7}R^{J_6}_{\phi_N}$ around $\phi_N = 0.0012$. In range 2, both $\varepsilon^{v_6}_{\phi_N}$ and $\varepsilon^{v_7}_{\phi_N}$ indicate that neither reaction 6 nor 7 has reached saturation. At the end

of range 3 $\varepsilon^{v_7}_{\phi_N}$ still has a significant positive value, while $\varepsilon^{v_6}_{\phi_N}$ has declined to nearly zero (indicating that reaction 6 is far from equilibrium and close to full saturation). For the control coefficients $C^{J_6}_{v_6}$ and $C^{J_6}_{v_7}$, the situation is somewhat reversed. Here, while having significantly lower values than their corresponding elasticity coefficients for most ϕ_N-values, the decline in $C^{J_6}_{v_7}$ is much more dramatic than that of $C^{J_6}_{v_6}$. At $\phi_N = 0.2$, $C^{J_6}_{v_7}$ is nearly zero, while at the highest tested ϕ_N-value $C^{J_6}_{v_6}$ and $\varepsilon^{v_6}_{\phi_N}$ are nearly equal. These results indicate that the decline of the partial responses ${}^{v_6}R^{J_6}_{\phi_N}$ and ${}^{v_7}R^{J_6}_{\phi_N}$ at higher ϕ_N-values is mostly due to the decline in control of v_7 on J_6 for ${}^{v_7}R^{J_6}_{\phi_N}$ and the decline in elasticity of v_6 towards ϕ_N for ${}^{v_6}R^{J_6}_{\phi_N}$. For ${}^{v_3}R^{J_6}_{\phi_N}$, both $C^{J_6}_{v_3}$ and $\varepsilon^{v_3}_{\phi_N}$ contribute to the declining negative partial response coefficient. We saw, however, that the inflection point observed in ${}^{v_3}R^{J_6}_{\phi_N}$ at $\phi_N = 0.4$ (Figure 4.7) is due to the contribution of the control coefficient rather than that of the elasticity coefficient.

4.4.2 Regulatory connections via feedback and isoenzymes

The aspartate-derived amino acid synthesis pathway model in Figure 4.3 contains a number of features that allow for multiple routes of regulation. Three of the steps are catalysed by isozymes that are allosterically modified by a variety of pathway intermediates and the pathway has multiple branch points. The isoenzymes also differ in terms of their kinetic properties and therefore respond differently to changes in the concentrations of their effectors. In this section we explore the importance of various routes of regulation of one intermediate with its supply block and elucidate the roles the various isoenzymes play within these routes.

4.4.2.1 Routes of interaction through different allosteric feedbacks

The first branch point of the aspartate-derived amino acid synthesis pathway occurs at aspartate-semialdehyde (ASA), which is produced by aspartate-semialdehyde dehydrogenase (ASADH or reaction 5) and consumed by two separate metabolic branches. The first step of the branch that produces threonine (Thr), cysteine (Cys) and isoleucine (Ile) as end products is catalysed by two isoenzymes, homoserine dehydrogenase I and II (HSDHI and HSDHII; reactions 6 and 7). The first step in the branch that produces lysine (Lys) as an end product is catalysed by the isoenzymes dihydrodipicolinate synthase 1 and 2 (DHDPS1 and DHDPS2; reactions 14 and 15). Thr and Lys inhibit the isoenzymes catalysing the first step of their respective branches, as well as two of the four aspartate kinase isoenzymes catalysing the first step in the pathway: Thr inhibits aspartate kinase I and II (AKI and AKII; reactions 1 and 2) and Lys inhibits aspartate kinase 1 and 2 (AK1 and AK2; reactions 4 and 3). Demand blocks of ASA can be defined according to the four consuming enzymes, with fluxes J_{14} and J_{15} in the Lys-branch, and J_6 and J_7 in the Thr-branch. Alternatively, we can define the demand blocks

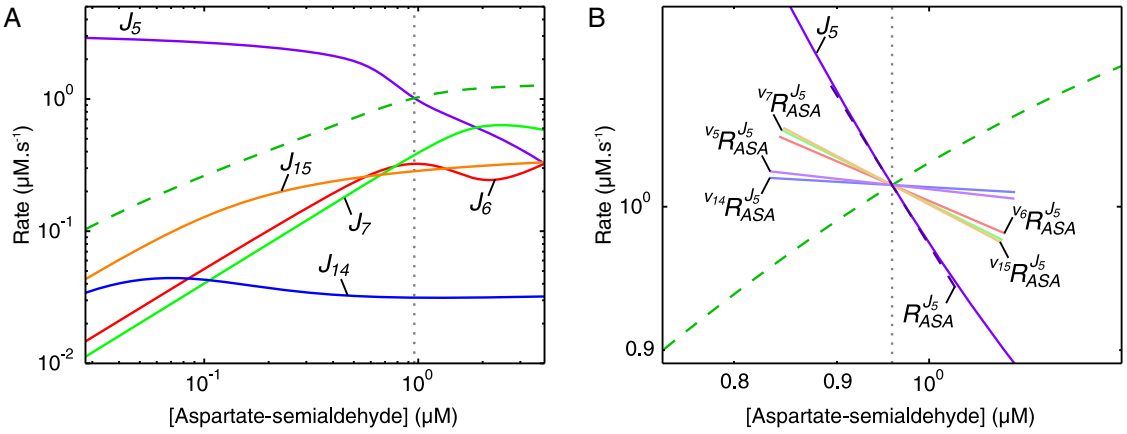


Figure 4.9: Rate characteristic plots of the reaction blocks of aspartate-semialdehyde in the aspartate metabolism model. **(A)** The fluxes of the demand blocks 6, 7, 14 and 15, and the supply block 5 of ASA. **(B)** The rate characteristic plot of the ASA-supply block 5 with partial and total response coefficients indicated as lines intersecting the J_5-curve at the steady-state value of ASA. Partial response coefficients (solid lines) indicate the relative contribution of each respective route of interaction of ASA with block 5 towards the total response coefficient (dashed line). The unlabelled dashed curve in both **(A)** and **(B)** represents the total demand and the steady-state value of ASA is indicated as a vertical dotted line (see Table 4.2).

according to the two separate metabolic branches where $J_6 + J_7 = J_8$ for the Thr branch and $J_{14} + J_{15} = J_{16}$ for the Lys branch. The rate characteristic plot shown in Figure 4.9A shows that at steady-state most of the flux proceeds towards Thr production with J_6 and J_7 being nearly equal. The flux towards Lys is carried mostly by block 15.

The inhibition of the AK isoenzymes by Thr and Lys enables changes in ASA to be transmitted to its supply block via these intermediates, which occur in the ASA-demand blocks. Partial response coefficients were used to determine the contribution of each of these routes of interaction to $R^{J_5}_{ASA}$ and therefore to quantify their importance in regulating J_5 (Figure 4.9B). Because of the very low degree of control of reaction 5 over its own flux, the direct interaction of ASA with block 5 via reaction 5 makes the second smallest contribution towards the total response despite its relatively high elasticity towards ASA (see Table S1 of Additional file 1). Instead the interactions of ASA with block 6, block 7 and block 15 contribute the most towards the observed total response. The enzymes catalysing the first steps of these blocks (reactions 6, 7 and 15) have lower sensitivities towards ASA than reaction 5, but much more control over J_5. The isoenzymes in each branch have practically identical elasticity coefficients towards ASA at the steady-state, therefore the difference in responses between the individual reactions in each branch is due to the differences in the degree to which these reactions control J_5, as determined by the flux carried by each reaction.

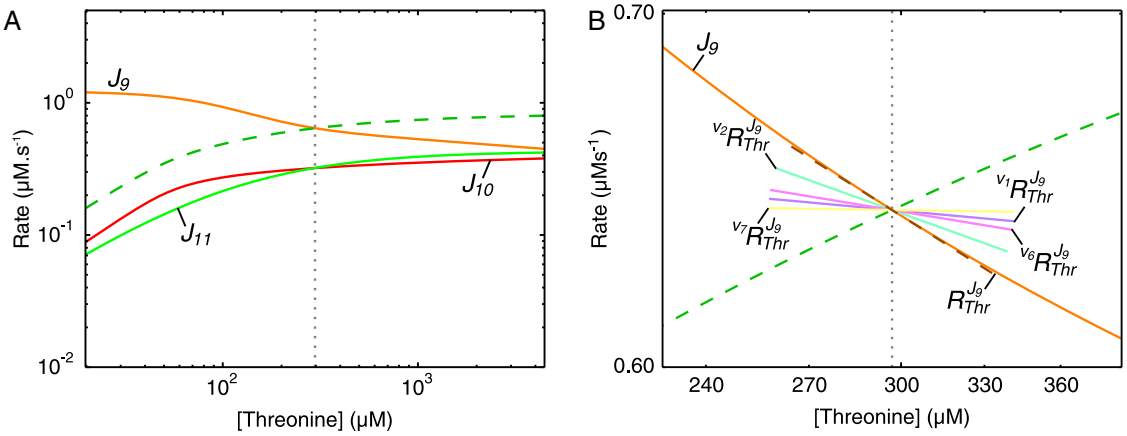


Figure 4.10: Rate characteristic plots of the reaction blocks of threonine in the aspartate metabolism model. **(A)** The fluxes of the demand blocks 10 and 11, and the supply block 9 of Thr. **(B)** The rate characteristic plot of the Thr-supply block 9 with partial and total response coefficients indicated as lines intersecting the J_9-curve at the steady-state value of Thr. Partial response coefficients (solid lines) indicate the relative contribution of each respective route of interaction of Thr with reaction block 9 towards the total response coefficient (dashed line). Note that ${}^{\nu_9}R^{J_9}_{Thr}$, ${}^{\nu_{10}}R^{J_9}_{Thr}$ and ${}^{\nu_{11}}R^{J_9}_{Thr}$ were omitted due to their zero contributions towards $R^{J_9}_{Thr}$. The unlabelled dashed curve in both **(A)** and **(B)** represents the total demand and the steady-state value of Thr is indicated as a vertical dotted line (see Table 4.2).

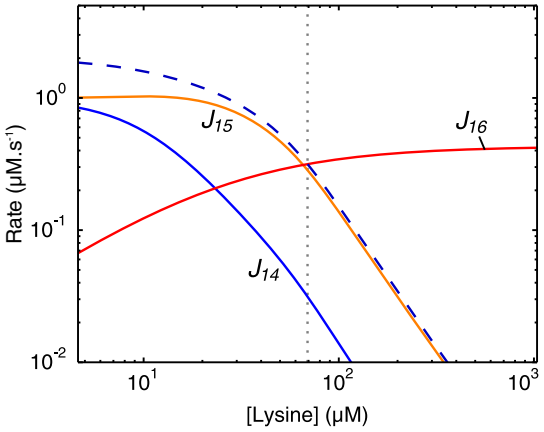


Figure 4.11: Rate characteristic plot showing the fluxes of the reaction blocks of lysine in aspartate metabolism. The demand block 16 and supply blocks 14 and 15 are indicated as solid lines. The unlabelled dashed curve represents the total supply. The steady-state value of Lys is indicated as a vertical dotted line (see Table 4.2).

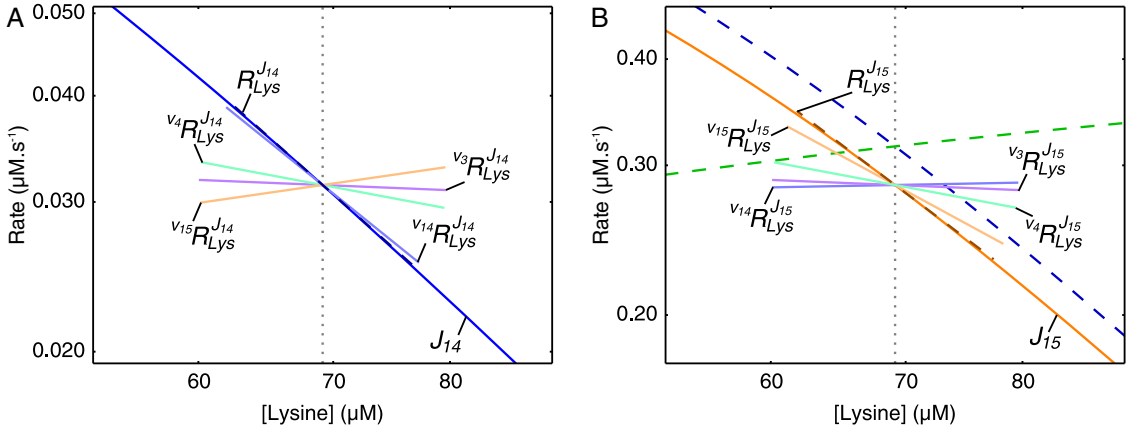


Figure 4.12: Rate characteristic plots of the supply blocks of lysine in the aspartate metabolism model. Partial and total response coefficients are indicated as lines intersecting the **(A)** J_{14}-curve and **(B)** J_{15}-curve at the steady-state value of Lys (see Table 4.2). Partial response coefficients (solid lines) indicate the relative contribution of each respective route of interaction of Lys with reaction blocks 14 and 15 towards the total responses (dashed lines) at the steady-state. The unlabelled dashed curves represent the total supply (blue) and total demand (green). The steady-state value of Lys is indicated as a vertical dotted line in both **(A)** and **(B)**.

While the above results quantify the importance of each of the routes of ASA supply block regulation and show that inhibition of reactions 1–4 by Thr and Lys plays an important regulatory role, we still have to quantify the amount of regulation that takes place via each of these four AK isoenzymes. This can be achieved by quantifying the contribution of the regulatory routes of the supply blocks of Thr and Lys and combining these results with the previous results.

The rate characteristic plot shown in Figure 4.10A illustrates the behaviour of the reaction blocks of Thr in response to changes in this metabolite's concentration. Here the Thr supply block (block 9), which ends with the enzyme threonine synthase (reaction 9), also encompasses the ASA supply block. Figure 4.10B shows the partial responses of J_9 towards Thr at the steady-state. It is clear that, as reaction 9 is insensitive towards its product Thr, the observed flux response is solely due to the inhibition of the upstream reactions 1, 2, 6 and 7 by Thr. In order to quantify the regulation of the ASA supply block, only the partial response coefficients of reaction block 9 towards Thr that represent routes passing through the ASA supply block, i.e., reactions 1 and 2, are of interest. We saw that despite only 1.8% and 9.4% of total flux respectively passing through reactions 1 and 2, and despite the resulting small degree of control these reactions have of over J_9, their high elasticities towards Thr cause both ${}^{v_1}R^{J_9}_{Thr}$ and ${}^{v_2}R^{J_9}_{Thr}$ to contribute significantly towards $R^{J_9}_{Thr}$. However, regardless of the specific contributions of these two routes in regulating J_9, the proportion of each of ${}^{v_1}R^{J_9}_{Thr}$

and ${}^{v_2}R^{J_9}_{Thr}$ to their total indicate the proportion of ASA supply block flux regulation taking place through reactions 1 and 2 (Figure 4.13A). Using the regulation of the ASA supply block by ASA via the route that passes through blocks 6 and 7 and subsequently through reaction 1 due to its inhibition by Thr as an example, the percentage regulation of this block taking place via reaction 1 (denoted by ${}^{v_1}\chi^{J_5}_{ASA}$) can be calculated as follows:

$$\begin{aligned} {}^{v_1}\chi^{J_5}_{ASA} &= \frac{{}^{v_6}R^{J_5}_{ASA} + {}^{v_7}R^{J_5}_{ASA}}{R^{J_5}_{ASA}} \times \frac{{}^{v_1}R^{J_9}_{Thr}}{{}^{v_1}R^{J_9}_{Thr} + {}^{v_2}R^{J_9}_{Thr}} \times 100 \\ &= \frac{(-0.276) + (-0.320)}{-1.043} \times \frac{-0.035}{(-0.035) + (-0.137)} \times 100 \\ &= 11.628\% \end{aligned} \tag{4.9}$$

Lys is produced by reactions 14 and 15 and therefore has two supply blocks. In this model the multi-step process of converting ASA to Lys was combined into a single step due to the irreversibility of reactions 14 and 15 [85]. At the steady-state, block 15 carries much more flux towards Lys than block 14, which, as we will see, affects the regulation of these blocks via each other (Figure 4.11). The partial responses of the two Lys supply blocks to Lys are shown in Figure 4.12. In contrast to the regulation of the Thr supply block where inhibition of the AK isoenzymes by Thr is the most important regulatory route, inhibition of the initial step of the Lys branch (reactions 14 and 15) by Lys is more important in eliciting a response in each of J_{14} and J_{15} than is inhibition of the AK isoenzymes. Interestingly, the inhibition of reaction 15 by Lys results in a positive response in J_{14} due to the attenuation of competition for the substrate ASA. The mirror effect of Lys-inhibition of J_{14} on J_{15} is also observed, but is smaller, as J_{14} carries much less flux towards Lys; reaction 14 therefore has less control over J_{15} than reaction 15 has over J_{14}. In the case of reactions 3 and 4, which are relevant to the routes of regulation of ASA with its supply block, we saw that reaction 4 contributes more towards the observed negative total response coefficients $R^{J_{14}}_{Lys}$ and $R^{J_{15}}_{Lys}$ than reaction 3 due to having significantly more control over J_{14} and J_{15} (see Table S1 of Additional file 1). The same technique used to elucidate the importance of reactions 1 and 2 in the regulation of the ASA supply block by ASA (as demonstrated in Equation 4.9) was used here to determine the importance of the routes of regulation involving reactions 3 and 4 (Figure 4.13A).

4.4.2.2 The effect of isoenzyme knockouts

The previous section showed the dissection of the routes of regulation of the ASA supply block rate by ASA for the wild type pathway. Using the same techniques, it was possible to gain insight into the regulation of this reaction block by ASA under alternative conditions. Here we performed the same analysis for knockout models of (1) AKI, (2) AKII , and (3)

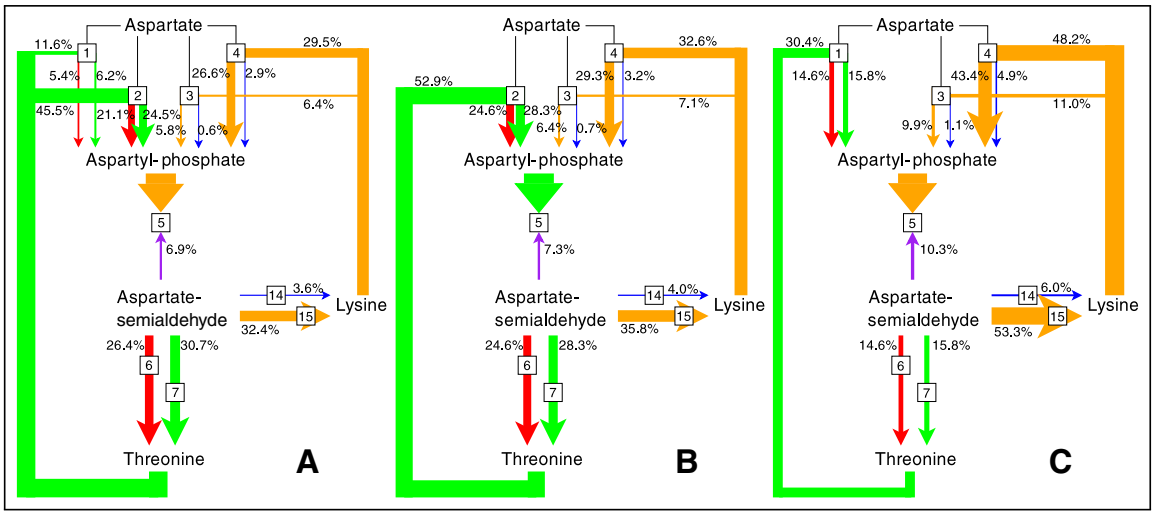


Figure 4.13: The importance of the various routes of regulation of ASA with its supply block. **(A)** The reference model. **(B)** AKI knockout. **(C)** AKII knockout. Knockouts in **(B)** and **(C)** were performed by respectively setting the concentrations of AKI and AKII to zero. All models were at steady-state (see Table S2 in Additional file 1).

both AKI and AKII. These knockouts, among others, were previously modelled by Curien *et al.* [203]. Steady-state analysis was performed as shown in Table S2 of Additional file 1 and the quantification of the importance of the various routes of regulation of J_5 by ASA is shown in Figures 4.13B and C.

Together, reactions 1 and 2 contribute only 12% of the total flux of the pathway at the reference steady-state (Table S2 of Additional file 1); however, as shown in the previous section, most of the regulation of the ASA supply block by ASA takes place via these two reactions. When taking all the AKs into consideration, reaction 2 is the most important and reaction 1 the third most important in terms of regulation. While the importance of the AK isoenzymes in terms of ASA supply flux regulation cannot be predicted by their flux contributions alone, there is nevertheless still a relationship between isoenzyme flux and regulatory importance because the degree of ASA supply flux control by the isoenzymes is a function of their relative flux contributions. Reaction 2 contributes 9.4% of the total flux with 45.5% of regulation taking place through it, while reaction 1 contributes 1.8% of total flux with only 11.6% of regulation taking place through this reaction (Figure 4.13A).

The knockout models highlighted once more the disconnection between flux and regulatory importance. The AKI knockout causes less than a 1% decrease in total flux (Table S2 of Additional file 1), with a concomitant decrease in regulation via the Thr inhibition route from 57.1% to 52.9% (Figure 4.13B). Here it was clear that while regulation via this route decreases, reaction 2 can compensate for the loss of reaction 1 by largely taking over its

Table 4.3: Analysis of the distribution of flux control between the supply and demand blocks of ASA. Total response coefficients of J_5, J_8 and J_{16} towards ASA at the steady-state and the distribution of flux control between the supply and demand blocks according to the principle illustrated in Equation 4.6 are presented. Control analysis was performed for the reference model and each of the knockouts with the concentration of the linking metabolite between supply and demand fixed at the steady-state value in each case (see Table S2 in Additional file 1).

	Reference	AKI knockout	AKII knockout	AKI, AKII knockout
$R^{J_5}_{ASA}$	-1.04	-0.95	-0.66	-0.46
$R^{J_8}_{ASA}$	0.52	0.53	0.59	0.61
$R^{J_{16}}_{ASA}$	0.14	0.14	0.14	0.15
$\mid R^{J_5}_{ASA} \,/\, R^{J_8}_{ASA} \mid$	2.01	1.80	1.13	0.76
$\mid R^{J_5}_{ASA} \,/\, R^{J_{16}}_{ASA} \mid$	7.56	6.85	4.60	3.17

regulatory role. On the other hand, a knockout of reaction 2 causes a 3.7% (Table S2 of Additional file 1) decrease in flux, which subsequently causes regulation via Thr inhibition to drop to 30.4% (Figure 4.13C). In spite of the regulatory importance of reaction 1 increasing by 162.1% compared to the wild-type (in contrast to a 83.3% increase in flux through this reaction), it cannot fully compensate for the loss of reaction 2. For all three knockout models, regulation is diverted mostly towards the Lys branch with 86.1% of regulation occurring via this branch for the double knockout (not shown), once more indicating the relatively low importance of the direct route of interaction of ASA with its supply block.

In addition to affecting the relative importance of the routes of regulation of the ASA supply block, the isoenzyme knockouts also affect the magnitude of the ASA supply and demand block responses (Table 4.3). There is a decrease in $R^{J_5}_{ASA}$ for each knockout model, with this effect being the least pronounced for the AKI knockout and the most pronounced for the double knockout. There is an increase in the Thr branch response ($R^{J_8}_{ASA}$) for the three knockouts, while the Lys branch response ($R^{J_{16}}_{ASA}$) remains relatively unchanged. These changes to the ratio of supply to demand response lead to changes in the functions of the reaction blocks in terms of flux control and ASA homoeostasis. At the reference state, both ASA demand branches have more control over their flux than the supply block, as indicated by the values of $|R^{J_5}_{ASA}/R^{J_8}_{ASA}|$ and $|R^{J_5}_{ASA}/R^{J_{16}}_{ASA}| > 1$ (Table 4.3). The decrease in $R^{J_5}_{ASA}$ observed for the knockout models causes a decrease in the ratio of the supply response to the demand response, indicating that flux control shifts towards the supply block. For the double knockout, this shift causes a reversal of roles in terms of flux control where the supply block has more

control of flux than the demand block.

4.5 Discussion

In this paper we set out to investigate multiple routes of interaction between reaction blocks in metabolic systems using the framework of generalised supply-demand analysis. This allowed us to identify routes of regulation and quantify the contributions of various reaction blocks towards metabolic behaviours such as flux responses. The two models investigated were chosen on the basis of the mechanism whereby multiple regulatory routes are mediated: moiety-conserved cycles connect different metabolic branches in the pyruvate branch model, while allosteric inhibition of various isoenzymes allows for communication between separate ends of the pathway in the aspartate model. In both cases, our analysis provided insight into how different routes of interaction contribute to the overall behaviour of the systems.

In the model of pyruvate metabolism in *Lactococcus lactis* multiple routes of interaction between reaction blocks are brought about by the moiety-conserved cycles of ATP/ADP, $NADH/NAD^+$ and acetyl-CoA/CoA. The regulation of this pathway in *L. lactis* is relatively well understood due to the extensive study of this industrially significant organism [204]. Moreover, a number of newer and more extensive models for *L. lactis* central carbon metabolism exist [192, 205–207], but most only include a simplified pyruvate branch metabolism. The most recently published model [207], for instance, lumps certain reactions such as acetaldehyde dehydrogenase and alcohol dehydrogenase, and phosphotransacetylase and acetate kinase together. Since our primary aim was to delineate and quantify regulatory routes of interaction, we chose to investigate the original model by Hoefnagel *et al.* [86], because it incorporates in the most detail the multiple interactions between the different branches of pyruvate metabolism due to the presence of moiety-conserved cycles. While an extended version of this model exists [192], which also incorporates glycolytic reactions in detail, the simpler version was chosen, as both have identical representations of the pyruvate branch metabolism and the extended model has not been published in full detail. We opted for an exploratory approach, focusing on the application of GSDA to this model in order to extract information about the effects of ATP/ADP and $NADH/NAD^+$ on various reaction blocks via multiple routes. This generated results that matched previous observations, but allowed us to offer a novel quantitative explanation.

The most striking result was that moiety ratios far from the steady-state caused unexpected flux responses in two reaction blocks: The ATP/ADP-producing flux J_5 responded positively to low ATP/ADP-values, while the $NADH/NAD^+$-consuming flux J_6 responded negatively to high $NADH/NAD^+$-values. The reactions, respectively catalysing the last and first

reactions in these reaction blocks, are not product activated or substrate inhibited in the model [86], therefore the observed flux responses had to originate from multiple routes of regulation. This was confirmed by utilising partial response coefficients to quantify the relative contribution of each route of interaction towards the total response of the two reaction blocks. The dominant route of regulation of its supply flux J_5 by all ATP/ADP-values below 30 was via the upstream lumped glycolysis reaction (v_1), rather than the direct route via the ATP-producing enzyme acetate kinase (v_5). This included both the steady-state value of ATP/ADP and the ATP/ADP-range where J_5 had a positive response. This is most probably an incomplete picture of regulation by ATP/ADP, as the inhibition by ATP and ADP of *L. lactis* enzymes such as lactate dehydrogenase, alcohol dehydrogenase and glyceraldehyde 3-phosphate dehydrogenase [208, 209] was not included in the model. It is conceivable that these additional routes of interaction could significantly affect the flux responses investigated here. It is, however, premature to speculate on any specifics without performing further work, due to the added complexity accompanying these interactions. Nevertheless, these results illustrate how a few routes of interaction can bring about unintuitive, non-monotonic flux responses, and how the different routes can be quantified in terms of their contributions towards these responses.

We found that while the direct route of interaction of NADH/NAD$^+$ with reaction block 6 via acetaldehyde dehydrogenase (v_6) mostly determined the behaviour of J_6 at steady-state, the interaction via pyruvate dehydrogenase (v_3) dominated at higher NADH/NAD$^+$-values, thereby causing a decrease in J_6. In spite of the limitations of this model, this corresponds well with the previously established role of redox balance in regulating pyruvate flux distribution, where low NADH/NAD$^+$-values are associated with mixed-acid fermentation and higher values with homolactic fermentation [210–213]. While a high sensitivity of lactate dehydrogenase towards NADH/NAD$^+$ [210] was not observed here, the reduction of flux towards acetyl-CoA (J_3) by inhibition of pyruvate dehydrogenase, and therefore also the ethanol flux (J_6), in response to the increase in NADH/NAD$^+$ was indeed observed [210–213]. Due to the structure of this pathway, one may conclude that reduced J_3 should lead to a reduction in J_6, but in reality matters are not that simple. While J_6 and J_3 did decrease concomitantly at higher NADH/NAD$^+$-values, there were also values for which J_6 increased while J_3 decreased. The observed J_6-response towards NADH/NAD$^+$ was shown to be a combination of complementary and competing non-monotonic effects that varied in importance with the value of the moiety ratio, thereby highlighting the utility of a model analysis tool such as GSDA for providing quantitative explanations for observed system behaviour.

Unsurprisingly, NADH/NAD$^+$ has been shown to determine pyruvate flux distribution in other organisms, such as *Saccharomyces cerevisiae* [214] and *Escherichia coli* [215, 216], in

a similar manner to *L. lactis* [211]. For these organisms, similar analyses could improve our understanding not only of their individual metabolisms, but also of pyruvate distribution in general. Furthermore, in addition to the role of NADH/NAD^+ in energy metabolism, NAD^+ and $NADP^+$ also play roles ranging from antioxidation to telomere metabolism as discussed in a comprehensive review by Ying [217]. While the approach used here may not be appropriate for the study of every role of NADH/NAD^+, its application could shed light on the specific regulatory role of NADH/NAD^+ in other pathways.

The second model investigated describes aspartate-derived amino acid synthesis in *Arabidopsis thaliana*. Here we focussed on the regulation of the ASA supply block by ASA itself. This reaction block was of special interest as its first step is catalysed by four AK isoenzymes, two of which are inhibited by Thr, and the other two inhibited by Lys. Each of these inhibitors is produced by a separate metabolic branch, with two isoenzymes catalysing the initial step of each ASA-consuming branch. These features enable ASA to communicate with its supply block via multiple routes.

Our results show that the majority of regulation of the ASA supply block did not occur via the interaction of ASA with its producing reaction, but rather by interaction with its demand blocks, which in turn affected the concentrations of the AK inhibitors Lys and Thr. One intuitively expects that regulation should occur via the shortest route, especially when taking into account the relatively high sensitivity of aspartate-semialdehyde dehydrogenase towards ASA at the steady-state in this system. The most unexpected result is the apparent importance of AKI and AKII in the ASA supply block regulation, in spite of their low contribution towards total flux. Previously, Curien *et al.* [203] analysed knockout simulations of this model and showed that AKI and AKII could compensate for the loss of AK1 and AK2 in terms of flux, thereby providing redundancy and confirming the idea that the role of isoenzymes is to provide robustness to the system [218, 219]. We, however, postulate that these reactions play an additional regulatory role which is largely decoupled from their function as carriers of flux, and that AKI and AKII provide robustness in terms of this role for each other. Our own knockout simulations of AKI and AKII showed that, while total flux remained largely unchanged for both knockouts (showing these reactions to be practically redundant in terms of flux), they were not redundant in terms of regulatory importance. While the loss of AKI could be compensated for by AKII, the reverse was not true, and a shift of regulation towards the Lys branch took place. However, in spite of the inadequacy of AKI as a substitute for AKII in terms of regulatory function and flux contribution, it was much more effective in emulating the former function than the latter. Furthermore, a double knockout of AKI and AKII decreased the total regulation of J_5 by ASA to less than 50% of the wild-type, shifting flux control from the demand block to, less optimally [14], the supply block. This means that

increases or decreases in Thr demand will no longer lead to effective regulation of the ASA flux.

The source of the flux responses of both models was investigated by separating the partial response coefficients into their control and elasticity components. In this way the flux response coefficients could be classified as originating predominantly from a local (i.e. enzyme) property or from a system property. We broadly classified a control or elasticity coefficient as having a dominant contribution towards the response coefficient in two different ways: either (1) the magnitude of one coefficient outweighs the contribution of the other, or (2) one coefficient changes in value over a parameter range while the other remains relatively constant; the varying coefficient therefore determines the change in response coefficient. In certain cases we found that the elasticity coefficients dominated the flux response (e.g., the large values of ${}^{v_7}R^{J_6}_{\phi_N}$ and ${}^{v_6}R^{J_6}_{\phi_N}$ at $\phi_N = 0.0012$ due to the huge values of $\varepsilon^{v_7}_{\phi_N}$ and $\varepsilon^{v_6}_{\phi_N}$ in the pyruvate model), while in other cases control coefficients dominated (e.g., the low value of ${}^{v_5}R^{J_5}_{ASA}$ at the steady-state due to the low value of $C^{J_5}_{v_5}$ in the aspartate model).

The work presented here reiterates the fact that metabolic systems can exhibit complex behaviour that cannot be predicted from simple inspection of the reaction network. Even when the network structure is considered together with enzyme-kinetic properties, in some cases understanding does not emerge intuitively. Furthermore, as the size and complexity of a system increases, so too does the variety of possible behaviours. Examples are the instances of apparent substrate inhibition and product activation in the pyruvate model (Figures 4.4A and 4.6A), where no such mechanisms exist on the enzyme level. Another example is in the aspartate model, where the seemingly predictable negative response of J_5 towards ASA is not due to product inhibition of ASADH, but rather due to upstream inhibition of the aspartate kinase isoenzymes by inhibitors downstream from ASA (Figure 4.9). The phenomena in both these cases stem from the existence of multiple routes of interaction between metabolites and reactions and were only brought to light through simulation and analysis. We could not only demonstrate unintuitive behaviour, but also quantify the contribution of the different routes of interaction towards bringing about this behaviour.

It is possible to analyse regulation at a deeper level by analysing the control coefficients that form part of the partial response coefficients in term of so-called control patterns [17]. A control pattern can be understood as a 'chain of local effects' that propagates through a metabolic pathway following a perturbation in a pathway parameter such as enzyme concentration. Each control pattern is a scaled product of elasticity coefficients, and each control coefficient is a sum of control patterns. Going even deeper, it is possible to partition the constituent elasticity coefficients into additive kinetic and thermodynamic terms [16]. The fact that control coefficients are complex functions of elasticity coefficients is also the rea-

son why certain control coefficients, such as $C^{J_6}_{v_3}$ and $C^{J_6}_{v_6}$ in the pyruvate model, responded non-monotonically to changing parameters, whereas the elasticities themselves responded monotonically. In this study there was only one situation where we could unambiguously relate an observed flux response to one of the terms in an elasticity coefficient: in the pyruvate model the infinite elasticities of reactions 6 and 7 towards ϕ_N at $\phi_N = 0.0012$ were due to these reactions being near equilibrium, a situation where for any reaction the thermodynamic term determines the value of its substrate and product elasticity coefficients. This observation was therefore only possible due to infinite elasticity coefficients being a very obvious and well-known sign of a reaction near equilibrium. To fully understand the pathways investigated here in terms of control patterns or in terms of the thermodynamics and kinetics of the pathway enzymes will require further analysis.

Both pathways studied here have potential biotechnological and industrial applications. *L. lactis* is an important organism in the dairy industry where the desirable products of pyruvate metabolism, such as diacetyl and acetaldehyde, are not always produced in equally desirable quantities [86, 210, 212]. Modification of *L. lactis* to increase these products is therefore an appealing prospect. While *A. thaliana* itself is not industrially important, it is used as a model organism for plant species in general. Here the modification of the aspartate-derived amino acid synthesis pathway to increase the production of the essential amino acids threonine and lysine could lead to the development of crops with increased nutritional value [220, 221]. However, the development of rational metabolic engineering strategies to leverage the metabolisms of these organisms requires a detailed understanding of their function. Application of the methods demonstrated in this paper can act as a stepping stone towards the development of such strategies by providing additional insights into mechanisms of metabolic regulation.

4.6 Conclusions

The regulation of the supply and demand blocks of a specific intermediate *by the intermediate itself* becomes convoluted when these reaction blocks can also interact through other intermediates, and not only through the linking intermediate. Generalised supply-demand analysis is a framework that allows for the identification of regulatory features of a metabolic pathway, one of which is the quantitative relative contribution of multiple routes of regulation of supply or demand blocks by the intermediate that links them.

Here we have demonstrated the use of generalised supply-demand analysis in disentangling various routes of regulation in a model of pyruvate metabolism where the involvement of the conserved moieties ATP/ADP and NADH/NAD^+ in multiple reactions caused counter-

intuitive responses in the fluxes of their producing and consuming blocks, and a model of aspartate metabolism where aspartate-semialdehyde could communicate with its supply block via multiple branching routes that were enabled by allosteric effectors and isoenzymes. Our findings showed that indirect routes of interaction between an intermediate and a reaction block can play a more significant role than the direct route.

We also demonstrated the utility of using a variety of analytic techniques in the further analysis of metabolic models. Both models provided novel results in spite of their having been studied by their original authors in the past [85, 86]. Further analysis with complementary tools such as control-pattern analysis would allow us to shed light on the source of the observed metabolic control in terms of chains of local effects [17, 23] and enzyme sensitivities in terms of thermodynamic and kinetic contributions [16]; computational implementations of these tools are currently in development.

Availability of supporting data

The original SBML versions of the models used in this paper can be found online in the BioModels Database [159] under the unique BioModels IDs BIOMD0000000017 and BIOMD0000000212 for the pyruvate branch and the aspartate-derived amino acid synthesis pathways, respectively. The PySCeS MDL and SBML versions of these models, together with a script to recreate the results presented here, are attached as "Additional Files" (see below). PySCeS MDL files of the models were obtained as described under "Methods".

Competing interests

The authors declare that they have no competing interests.

Author's contributions

CDC and JMR conceived the study. CDC wrote the modelling, GSDA and MCA scripts, analysed and interpreted the data and drafted the manuscript. JMR and JHSH supervised the project, helped with interpretation of data and critically revised the manuscript. All authors approved the final version of the manuscript.

Acknowledgements

This work is based upon research supported by the National Research Foundation (NRF) of South Africa. CDC is the recipient of an NRF PhD Innovation scholarship.

Additional Files

Additional file 1 — Hoefnagel_moiety_ratio.psc

PySCeS model descriptor language file for the pyruvate model, as required by the IPython notebook provided as Additional file 6 (see instructions within notebook).

Additional file 2 — Curien.psc

PySCeS model descriptor language file for the aspartate model, as required by the IPython notebook provided as Additional file 6 (see instructions within notebook). Software needed for Additional files 1 and 2: PySCeS (see http://pysces.sourceforge.net/docs/userguide.html).

Additional file 3 — Hoefnagel_moiety_ratio.xml

SBML [158] version of the pyruvate model.

Additional file 4 — Curien.xml

SBML [158] version of the aspartate model.

Additional files 3 and 4 have been tested with the following simulators:

- PySCeS (http://pysces.sourceforge.net/docs/userguide.html)
- JWSOnline (http://jjj.biochem.sun.ac.za/)
- Copasi (http://www.copasi.org)

Other SBML-compliant simulators (http://sbml.org/SBML_Software_Guide/SBML_Software_Matrix) may also work.

Additional file 5 — Results.ipynb

An IPython notebook file in which the results shown in Figures 4–13, Table 3 and Additional file 1 are generated using PySCeS and PySCeSToolbox. This notebook requires the model

files provided in Additional files 2 and 3. Software needed: IPython notebook with PySCeS and PySCeSToolbox as requirements (see https://github.com/PySCeS/PyscesToolbox for full requirements and installation instructions).

Additional file 6 — Additional tables.pdf

A pdf document containing Table S1 (metabolic control analysis) and Table S2 (steady-state analysis) for the aspartate metabolism model[2].

[2]These tables are included in Appendix B

Chapter 5

Delving Deeper: Relating the Behaviour of a Metabolic System to the Properties of its Components

5.1 Introduction

Metabolic systems are highly complex and interconnected networks consisting of numerous functional molecular components. These systems are particularly exemplary of the phenomenon of emergence, as the behaviour at the level of the system rarely relies on any single component, but rather arises from the unique properties and non-linear interactions of all its components. Unfortunately, this means that understanding these systems on a mechanistic level is a particularly challenging task as it requires quantitative knowledge of all components together with their interactions. In this regard, kinetic models of metabolic systems are essential tools [55, 66], as they potentially allow for quantitative insight into the behaviour of the systems they represent in a way that is otherwise impossible to achieve with even the most meticulous laboratory techniques. However, metabolic models are not, in themselves, sufficient to achieve the sought-after mechanistic understanding. To this end, there is a need for theory and tools that allow for systematic and quantitative investigation into the emergent properties of the system.

The framework of metabolic control analysis (MCA) is one such tool that aims to explain the behaviour of a system in terms of its components [8, 9]. This framework allows for the quantification of the sensitivity of system variables (such as fluxes and steady-state concentrations) towards perturbations in the reactions of a metabolic system. Additionally, it allows for these system variable sensitivities to be understood in terms of the network topology and

the pathway components through the application of the summation and connectivity theorems of MCA. The theory and methods that constitute MCA have been extensively expanded upon since its initial conception [11–13], and methods have been developed that generalise the summation and connectivity theorems [8]. However, MCA is frequently applied in order to understand the control of certain key metabolic variables with the end goal of metabolic engineering [36, 222], without considering how control is brought about. In other words, the question of explaining emergence is often ignored in favour of more practical goals. While favouring a more pragmatic approach is certainly understandable, it leaves much to be desired in terms of a deeper understanding of the control properties of metabolic systems. Moreover, deeper insight could also possibly aid in the development of more effective engineering strategies.

The thermodynamics of metabolic systems can also provide insight into their functioning. In structural metabolic models for flux balance analysis, for instance, the thermodynamics of a system is one of the criteria used to constrain the possible solution space [223]. In kinetic models, the distance of a reaction from equilibrium can indicate whether an enzyme-catalysed reaction is predominantly controlled by the properties of the enzyme itself, or by the intrinsic mass action effect [12, 16, 25, 136]. However, in the past, the practical application and utility of this idea has been limited due to an imprecise delineation between what can be considered near-equilibrium and far-from-equilibrium [12, 16]. Additionally, even if distance from equilibrium is precisely defined, it does not give complete insight into the relative importance of the effect enzyme binding vs. that of mass action [16].

One organism in which the systems biology approach has successfully yielded new insight is *Lactococcus lactis* (as reviewed in [224]). Much work has specifically gone into understanding glycolysis in this organism, as is evident from the variety of published models that describe this system [86, 192, 205, 207]. However, the exact mechanism behind the switch between mixed-acid fermentation and the lower ATP-yielding homolactic fermentation is yet to be uncovered [224]. Previous work in this regard has pointed to redox balance as playing an important role [210–214]. As part of a larger study that focussed on the quantification of regulatory routes in metabolic models [31] (Chapter 4), we utilised generalised supply-demand analysis [15] to uncover the effect of the redox balance on the different metabolic branches of pyruvate metabolism of *Lactococcus lactis* using a previously published metabolic model [86]. In this study, an increase in $NADH/NAD^+$ was shown to decrease flux towards acetaldehyde and ethanol, mirroring past experimental [210, 212–214] and FBA modelling [211] findings. This phenomenon was shown to originate predominantly from the interaction of $NADH/NAD^+$ with pyruvate dehydrogenase. Additionally, we quantified this effect in terms of control and elasticity coefficients, thereby distinguishing between the contribution

of systemic and local effects on producing the observed flux response.

In this chapter we build upon this previous work by considering the origin of the control and elasticity coefficients that constitute the response coefficient of the flux through acetaldehyde dehydrogenase in *L. lactis*. Two lines of investigation are followed. In the first we consider expressions of control in terms of elasticity coefficients and fluxes [23]. These symbolic expressions are examined within the framework of control-pattern analysis [17], which dissects them into physical chains of local effects, or control patterns, that show how control emerges from the interactions between the system components. By identifying, quantifying, and comparing common motifs within control patterns we determine the importance of different chains of effects over a range of NADH/NAD^+ values, thus explaining metabolic control in physiological terms. Secondly, we consider the behaviour of the enzyme-catalysed reactions themselves. Here we differentiate between the effect of enzyme binding and that of mass-action on the control of the reactions. Finally, we relate these results to one another and expand our investigation to the control patterns, showing how the properties of the individual reactions lead to the observed control profile. In this way we not only explore the properties of the system in question, but also attempt to demonstrate general principles for understanding metabolic systems within the context of these frameworks.

5.2 Methods

5.2.1 Metabolic control analysis

Metabolic control analysis (MCA) is a framework for quantifying the control properties of a steady-state metabolic system in terms of the responses of its fluxes and metabolite concentrations towards perturbations in the rates of its reactions [8, 9]. MCA also relates these sensitivities to the properties of the system components through the use of the connectivity and summation properties [8]. Below we briefly describe the fundamental coefficients of MCA and define their relationships.

Elasticity coefficients describe the sensitivities of isolated reactions towards changes in their parameters and the concentrations of their substrates, products, or direct effectors. For a reaction i affected by a metabolite X, the elasticity coefficient is defined as a ratio of relative change of the reaction rate v_i to the relative change of the concentration x:

$$\varepsilon_x^{v_i} = \frac{\partial \ln v_i}{\partial \ln x} \tag{5.1}$$

Control coefficients describe the sensitivities of system variables of a metabolic pathway, such as flux or steady-state metabolite concentrations, towards changes in the local rates of

its reactions. These sensitivities, unlike elasticity coefficients, are a function of the complete system and depend on the system topology and component properties and are thus defined using total derivatives. Control coefficients are defined as a ratio of relative change of any system variable y to the relative change of a reaction rate v_i:

$$C_i^y = \frac{d \ln y}{d \ln v_i} \tag{5.2}$$

The summation property describes the distribution of control between the reactions within a metabolic pathway. This property applies to all metabolic systems and states that the sum of the control coefficients of all reactions on any particular flux is equal to 1. For the two step system with reactions 1 and 2:

$$S \overset{1}{\rightleftharpoons} X \overset{2}{\rightleftharpoons} P \tag{5.3}$$

the flux summation property is defined as:

$$C_1^J + C_2^J = 1 \tag{5.4}$$

The connectivity theorem describes the relationship between control and elasticity coefficients. The flux connectivity states that in a system where a metabolite affects multiple reactions, the sum of the products of each of the control coefficients of a particular flux towards these reactions multiplied by the corresponding elasticity coefficients of these reactions towards the metabolite is equal to 0. For our two step system (Equation 5.3) where both reactions are affected by x, it can be expressed as:

$$C_1^J \varepsilon_x^{v_1} + C_2^J \varepsilon_x^{v_2} = 0 \tag{5.5}$$

Solving these simultaneous equations produces expressions for each of the two control coefficients C_1^J and C_2^J in terms of the elasticity coefficients:

$$C_1^J = \frac{\varepsilon_x^{v_2}}{\varepsilon_x^{v_2} - \varepsilon_x^{v_1}} \quad \text{and} \quad C_2^J = \frac{-\varepsilon_x^{v_1}}{\varepsilon_x^{v_2} - \varepsilon_x^{v_1}} \tag{5.6}$$

The relationship between control and elasticity coefficients expressed above can also be obtained through one of the various matrix-based formalisations of MCA [19–21, 56, 57, 98–110]. One such method, previously described in [57], combines the summation and connectivity properties into a generalised matrix form called the control matrix equation. Here a matrix of independent control coefficients, $\mathbf{C^i}$, and a matrix expressing structural properties and elasticity coefficients, $\mathbf{E}$, have the relationship:

$$\mathbf{C^i E} = \mathbf{I} \tag{5.7}$$

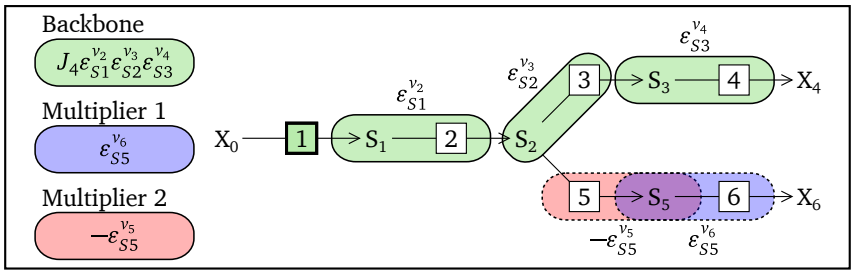


Figure 5.1: Control patterns for a 6 step branched pathway. The backbone and multiplier patterns for two control patterns of $C_1^{J_1}$ are depicted. Green bubbles indicate the backbone pattern, while red and blue bubbles each indicate a multiplier pattern. Each multiplier-backbone combination represents a single control pattern. The chain of local effects for the backbone originates from the perturbation of reaction 1 which affects the concentration of S_1, in turn affecting reaction 2, and so forth, until reaching reaction 4. An increase in the concentration of the enzyme catalysing reaction 1 will therefore cause a chain of positive effects, each playing a small role in determining the sensitivity of J_1 towards v_1.

This means that $\mathbf{C^i}$ can be calculated by the inversion of $\mathbf{E}$. In the case of a symbolic inversion of $\mathbf{E}$ [23], expressions similar to those shown in Equation 5.6 are obtained.

These expressions of control also translate into physical concepts. Each term of the numerator of a control coefficient expression represents a chain of local effects that radiates from the modulated reaction as demonstrated in Fig. 5.1 [17]. These chains of local effects, otherwise known as control patterns, describe the different paths through which a perturbation can propagate towards the controlled system variable, with each unique path potentially carrying a different weight towards the entire control coefficient. The contribution of a control pattern towards its control coefficient is quantified as its percentage absolute contribution towards the sum of the absolute values of all the control patterns of the control coefficient. This metric is used instead of a conventional percentage to account for cases where control patterns have different signs, which could lead to the contribution of a single pattern being more than 100%.

Control patterns can be further subdivided into smaller backbone and multiplier patterns [118]. In the case of flux control coefficients, a backbone pattern is defined as an uninterrupted chain of reactions that links two terminal metabolites and passes through the flux being controlled, known as the reference flux. Multiplier patterns are chains of reactions that are not directly linked with a backbone pattern, and occur only in the case of branched pathways. Different multiplier patterns can therefore be combined with a single backbone to form different control patterns. Conversely, the same multipliers can also be found among patterns with different backbones.

5.2.2 Regulation by enzymes

One definition of regulation, as it pertains to metabolic systems, is *the alteration of reaction properties to augment or counteract the mass-action trend in a network of reactions* [25]. One way that the mass-action trend can be counteracted is through the activity of enzymes, with higher potential for regulation being achieved far from equilibrium. Therefore, in order to quantify the regulatory effect of enzymes in a system, it is necessary to be able to determine the distance from equilibrium of a reaction, and to be able to distinguish between the effect of mass action and enzyme binding [16].

Distance from equilibrium is given by the disequilibrium ratio Γ/K_{eq} (or ρ), with kinetic control in the forward direction indicated by $\rho \leq 0.1$, thermodynamic control in the forward direction indicated by $\rho \geq 0.9$, and a combination of kinetic and thermodynamic control indicated by $0.1 < \rho < 0.9$ [16].

Determining the effects of binding and mass action requires a rate equation to be cast in a form which separates these two functions and partially differentiating its logarithmic form with respect to the logarithm of the substrates or products [16, 25]. This process yields two elasticity coefficients, one which quantifies the effect of the substrate or product binding on the reaction rate, and the other the effect of mass action. As an example, a Michaelis-Menten rate equation with substrate S and product P is expressed in a way that makes the rate contributions due to rate capacity (v_{cap}), binding (Θ), and mass action (v_{ma}) explicit as factors arranged from left to right:

$$v = \underbrace{\frac{V_f}{K_s}}_{v_{cap}} \times \underbrace{\frac{1}{1+\frac{s}{K_s}+\frac{p}{K_p}}}_{\Theta} \times \underbrace{\left(s-\frac{p}{K_{eq}}\right)}_{v_{ma}} \tag{5.8}$$

Partial differential of the natural log form of Equation 5.8 with respect to $\ln s$ yields binding (ε_s^{Θ}) and mass-action ($\varepsilon_s^{v_{ma}}$) elasticity expressions, which add up to the total elasticity (ε_s^{v}) [16]:

$$\begin{aligned}\varepsilon_s^v &= 0 + \frac{-\frac{s}{K_s}}{1+\frac{s}{K_s}+\frac{p}{K_p}} + \frac{1}{1-\rho} \\ &= \varepsilon_s^{\Theta} + \varepsilon_s^{v_{ma}}\end{aligned} \tag{5.9}$$

Similar expressions can be defined for the product. With these separate elasticity coefficients we can quantify, and differentiate between, the effect of either a substrate or a product on the reaction rate though enzyme binding and the mass-action effect.

5.2.3 Software

Model simulations were performed using the Python Simulator for Cellular Systems (PySCeS) [91] within a Jupyter notebook [173]. Symbolic inversion of the **E** matrix (Equation 5.7) and subsequent identification and quantification of control patterns was performed by the SymCa [23] module of the PySCeSToolbox [31] add on package for PySCeS. Control patterns for each control coefficient were automatically numbered by SymCa starting from 001, and we used these assigned numbers as-is in the presented results. Importantly, SymCa was set to automatically replace zero-value elasticity coefficients with zeros. In other words, certain elasticity coefficients are never found within any control coefficient expressions as their value will always be zero (such as in the case of irreversible reactions).

The automatic identification of the Γ/K_{eq} terms of the reactions, subsequent automatic separation of rate equations into the binding and mass-action terms, and calculation of elasticity coefficients for these terms was performed by the ThermoKin module of PySCeSToolbox. Additional manipulation of symbolic expressions, data analysis, and visualisations were performed using the SymPy [175], NumPy [179] and Matplotlib [174] libraries for Python [157].

5.2.4 Model

As mentioned in the Introduction to this chapter, we revisited results obtained in a previous study of a model of pyruvate branch metabolism in *Lactococcus lactis* [31] (Chapter 4). This model, shown in Fig. 5.2, was originally constructed by Hoefnagel *et al.* [86], and we retrieved it from the JWS online model database [160] in the PySCeS model descriptor language (see http://pysces.sourceforge.net/docs/userguide.html).

Members of the of ATP/ADP, acetyl-CoA/CoA and NADH/NAD^+ moiety-conserved cycles were treated as ratios, respectively represented by the symbols ϕ_A, ϕ_C, and ϕ_N, in order to perform parameter scans of these conserved moieties without breaking moiety conservation. In this chapter, only the effect of a change in the value of ϕ_N is considered. The value of ϕ_N was thus fixed and varied between 0.0002 and 1.77 in order to generate Figs. 5.3, 5.4, 5.8, 5.9 and 5.10.

The value of ϕ_N was varied directly rather than modulating its demand by changing the activity of NADH oxidase for two reasons. Firstly, we wanted follow the same methodology used in our original study [31] (Chapter 4), so that we may explain the results obtained there. Secondly, we wanted to simplify the system for control pattern analysis. These considerations are discussed further in Section 5.3.6, where we perform a similar control pattern analysis for a range of ϕ_N values by varying the V_{max} of NADH oxidase in a version of the model where

ϕ_N is not fixed.

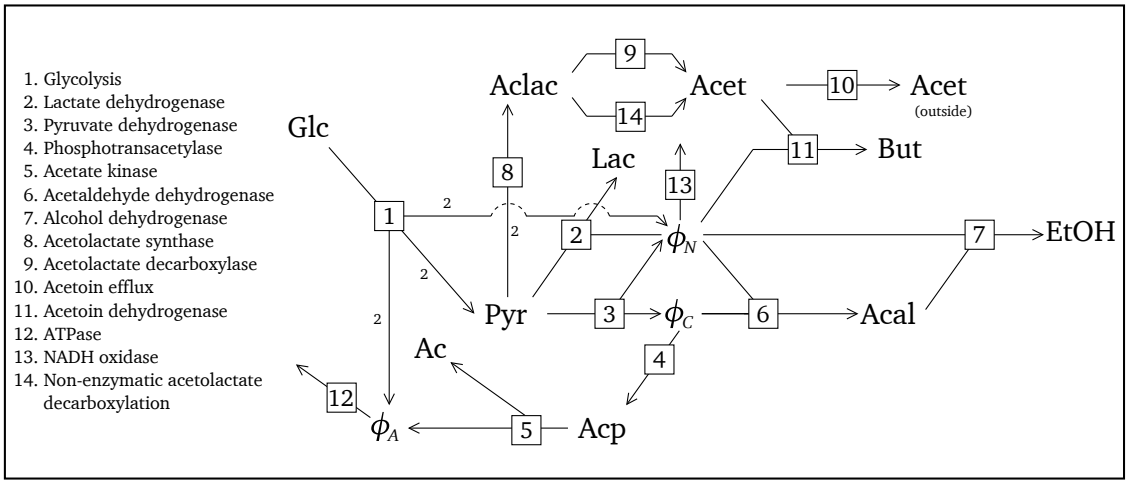


Figure 5.2: The pyruvate branch pathway as defined by Hoefnagel *et al.* [86]. Reactions are numbered according to the key. The stoichiometry of each reaction is 1 to 1, except for reaction 1 where $\mathrm{Glc} + 2\mathrm{ADP} + 2\mathrm{NAD}^+ \rightarrow 2\mathrm{Pyr} + 2\mathrm{ATP} + 2\mathrm{NADH}$ and reaction 8 where $2\mathrm{Pyr} \rightleftharpoons \mathrm{Aclac}$. Intermediates are abbreviated as follows: Ac: acetate; Acal: acetaldehyde; Acet: acetoin; Aclac: acetolactate; Acp: acetyl phosphate; Glc: glucose; Lac: lactate; But: 2,3-butanediol; Pyr: pyruvate; EtOH: ethanol.

5.3 Results and discussion

One of the main findings of our previous study [31] (Chapter 4) was that an increase in $\mathrm{NADH}/\mathrm{NAD}^+$ (referred to as ϕ_N henceforth) caused a decrease in the flux through the acetaldehyde dehydrogenase reaction block (J_6). By investigating the partial response coefficients of J_6 with respect to ϕ_N, we found that this effect was due to the interaction of ϕ_N with pyruvate dehydrogenase (reaction 3) as signified by the partial response coefficient ${}^{\nu_3}R^{J_6}_{\phi_N}$. Moreover, over the full range of tested ϕ_N values, three routes of regulation were found to be dominant, accounting for most of the observed response of J_6 towards ϕ_N. In addition to ${}^{\nu_3}R^{J_6}_{\phi_N}$, the other two dominant partial responses were ${}^{\nu_6}R^{J_6}_{\phi_N}$ and ${}^{\nu_7}R^{J_6}_{\phi_N}$. The three dominant partial response coefficients were also further divided into their component control and elasticity coefficients, showing how these components contributed to the overall observed response.

In the following sections we will continue this line of investigation by examining how $C^{J_6}_{\nu_6}$, $C^{J_6}_{\nu_7}$ and $C^{J_6}_{\nu_3}$ arise from the interactions between the various species and enzymes of the system, with particular focus on the latter control coefficient. In addition, we will investigate the corresponding elasticity coefficients in terms of the kinetic and thermodynamic properties

of their reactions. After considering the elasticity and control coefficients individually, we will combine these concepts, showing how reaction properties determine control. Finally, we compare the free-ϕ_N model with the fixed-ϕ_N model in terms of their control patterns.

5.3.1 Identification of dominant control patterns of $C^{J_6}_{v_3}$

Our investigation of $C^{J_6}_{v_3}$ commenced with the identification of its control patterns. The `SymCa` package identified 76 unique control patterns and generated expressions for each. While this is a significantly smaller number compared to the 226 patterns identified in the unclamped system (see Section 5.3.6), naively investigating the properties of each would still represent an unwieldy task. Therefore a subset representing the most important control patterns in terms of their contribution towards $C^{J_6}_{v_3}$ over the scanned ϕ_N range were chosen for further investigation. As our goal was to simplify our analysis by excluding all control patterns that contributed very little towards the overall control coefficient, we defined the important control patterns as the smallest selection of control patterns that could account for at least 70% of the absolute sum of all the control patterns for the complete ϕ_N range (see Section 5.2.1). Additionally, individual control patterns had to adhere to a certain minimum percentage absolute contribution for more than a certain percentage of the complete ϕ_N range (sampled on a logarithmic scale). The first cut-off value excluded the smallest control patterns, whereas the second excluded those only contributing more than the minimum of the first cut-off for a very small range of ϕ_N values. This selection process was automated by independently varying the two cut-off values between 1% and 15% in 1% increments, and selecting the smallest group that adhered to the 70% minimum. Admittedly, these criteria are arbitrary to some extent; however, they effectively reduced the number of control patterns we had to consider in a relatively unbiased manner while still accounting for the majority of the control.

As shown in Fig. 5.3A, the selection process yielded 11 “important” control patterns. These patterns make varying contributions towards $C^{J_6}_{v_3}$ depending on the value of ϕ_N, with different patterns being dominant within different ranges of ϕ_N values. Based on their contribution towards the control coefficient within a certain range and the general shape of their curves, control patterns were categorised into groups numbered 1–4 as indicated in Fig. 5.3. Each group of patterns therefore has a distinct ϕ_N range in which they are a responsible for the most of the observed value of $C^{J_6}_{v_3}$ with the exception of patterns within group 3 (CP063 and CP030), which only ever overlap with patterns within groups 2 and 4. Fig. 5.3B provides another perspective on the contribution of the 11 dominant control patterns towards the control coefficient by demonstrating how closely their sum follows the value of $C^{J_6}_{v_3}$.

These results demonstrate that, for certain conditions, some chains of physical effects are significantly more important for determining the control of one reaction over a flux than

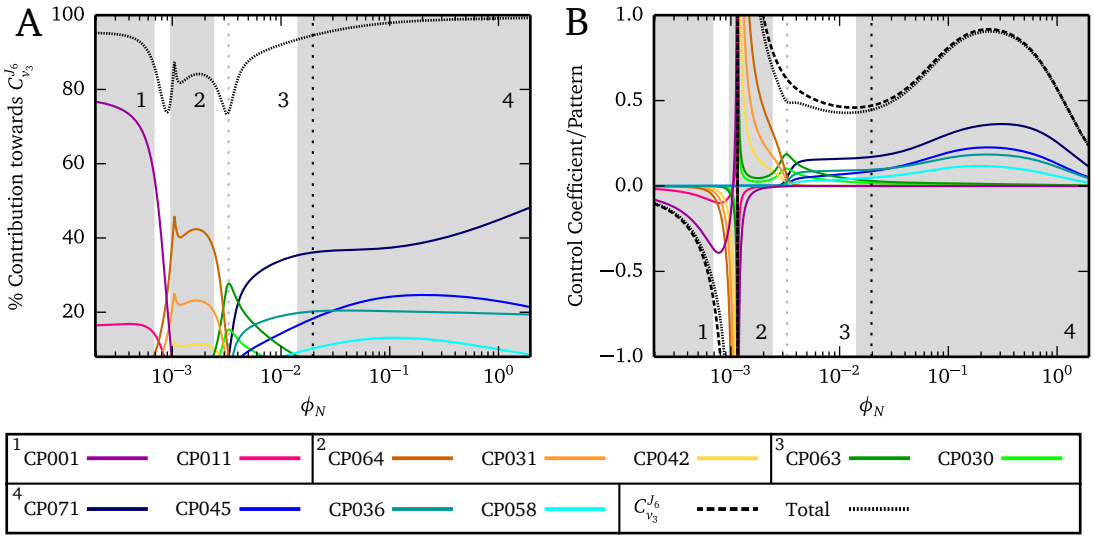


Figure 5.3: The most important control patterns of $C^{J_6}_{\nu_3}$ as functions of ϕ_N. Control patterns were chosen based on the criteria described in the text. **A** shows the absolute percentage contribution of the important patterns towards the absolute sum of all the control patterns of $C^{J_6}_{\nu_3}$. **B** shows these control patterns in relation to $C^{J_6}_{\nu_3}$ as well as their sum (the former is the upper dashed curve in region 3, whereas the latter is the dashed curve below it). Control patterns and $C^{J_6}_{\nu_3}$ are indicated in the key. The grey shaded regions indicate ranges of ϕ_N values which are dominated by a group of related control patterns, while unshaded regions indicate shared contribution by unrelated patterns. Groups are numbered 1–4 in both **A** and **B**, and control patterns belonging to a group are colour coded. The grey dotted vertical line indicates shared contribution between patterns in group 2 and 3 on the left hand side and between patterns in group 3 and 4 on the right hand side. The black dotted vertical line indicates the steady-state value of ϕ_N.

others. In our case, the vast majority of flux control could be attributed to a small selection of control patterns for a large range of ϕ_N values, with at most 6 patterns meeting the cut-off criteria for any ϕ_N value. Whether the remaining 85% of control patterns play any significant role under different conditions is completely dependent on what those conditions may be and how they affect the components of these patterns. Among the 11 important control patterns, different groups of patterns are more important than others depending on the value of ϕ_N, and there is a clear change of "regime" as the value of ϕ_N changes. However, without investigating the actual composition of these control patterns it is impossible to understand why certain patterns are more important than others and how their values, and degree of contribution towards $C^{J_6}_{\nu_3}$, are determined.

Table 5.1: Backbone and multiplier expressions of the control patterns of $C_{v_3}^{J_6}$. Backbone expressions are named A–F and multipliers are classified according to their relative position on the reaction scheme (Fig. 5.6), with B and T multipliers appearing on the bottom and top halves respectively. Each of the control pattern numerators of $C_{v_3}^{J_6}$ is a product of one of the backbones and either a B (for backbones B, D–F) or both B and T multipliers (for backbones A and C). Valid control pattern numerators do not consist of multiplier-backbone combinations with overlapping factors.

Factor	Expression
A	$-J_1J_3\varepsilon_{Pyr}^{v_1}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}$
B	$2J_3J_8J_9J_{10}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}\varepsilon_{Pyr}^{v_8}\varepsilon_{Aclac}^{v_9}\varepsilon_{Acet}^{v_{10}}$
C	$J_2J_3\varepsilon_{Pyr}^{v_2}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}$
D	$2J_3J_8J_{11}J_{14}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}\varepsilon_{Pyr}^{v_8}\varepsilon_{Acet}^{v_{11}}\varepsilon_{Aclac}^{v_{14}}$
E	$2J_3J_8J_{10}J_{14}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}\varepsilon_{Pyr}^{v_8}\varepsilon_{Acet}^{v_{10}}\varepsilon_{Aclac}^{v_{14}}$
F	$2J_3J_8J_9J_{11}\varepsilon_{\phi_C}^{v_6}\varepsilon_{Acal}^{v_7}\varepsilon_{Pyr}^{v_8}\varepsilon_{Aclac}^{v_9}\varepsilon_{Acet}^{v_{11}}$
B1	$-J_1\varepsilon_{\phi_A}^{v_1}\varepsilon_{Acp}^{v_5}$
B2	$-J_{12}\varepsilon_{Acp}^{v_4}\varepsilon_{\phi_A}^{v_{12}}$
B3	$J_{12}\varepsilon_{Acp}^{v_5}\varepsilon_{\phi_A}^{v_{12}}$
B4	$J_1\varepsilon_{\phi_A}^{v_1}\varepsilon_{Acp}^{v_4}$
B5	$2J_5\varepsilon_{Acp}^{v_4}\varepsilon_{\phi_A}^{v_5}$
B6	$J_5\varepsilon_{Acp}^{v_4}\varepsilon_{\phi_A}^{v_5}$
T1	$-J_8J_{10}\varepsilon_{Aclac}^{v_8}\varepsilon_{Acet}^{v_{10}}$
T2	$-J_8J_{11}\varepsilon_{Aclac}^{v_8}\varepsilon_{Acet}^{v_{11}}$
T3	$2J_{10}J_{14}\varepsilon_{Acet}^{v_{10}}\varepsilon_{Aclac}^{v_{14}}$
T4	$2J_9J_{10}\varepsilon_{Aclac}^{v_9}\varepsilon_{Acet}^{v_{10}}$
T5	$2J_{11}J_{14}\varepsilon_{Acet}^{v_{11}}\varepsilon_{Aclac}^{v_{14}}$
T6	$2J_9J_{11}\varepsilon_{Aclac}^{v_9}\varepsilon_{Acet}^{v_{11}}$
T7	$J_8J_9\varepsilon_{Aclac}^{v_8}\varepsilon_{Acet}^{v_9}$

5.3.2 Backbone and Multiplier Patterns of $C_{v_3}^{J_6}$

In order to investigate the source of the differences between the control patterns in terms of their values and contribution towards $C_{v_3}^{J_6}$, we identified the various backbone and multiplier patterns that they incorporate. These smaller patterns are useful for providing an intermediate level of abstraction between the full control patterns and their flux, concentration, and

Table 5.2: Numerator expressions of the dominant control patterns of $C_{v_3}^{J_6}$. The numerators of the control patterns highlighted in Fig. 5.3A are expressed in terms of their constituent backbone and multiplier factors (Table 5.1) and are separated into groups based on the ϕ_N ranges for which they are most important as in Fig. 5.3A. Control patterns are arranged in descending order of relative importance within their group.

Group	Control Pattern	Expression
1	CP001	$A \cdot B3 \cdot T1$
	CP011	$A \cdot B2 \cdot T1$
2	CP064	$B \cdot B3$
	CP031	$B \cdot B1$
	CP042	$B \cdot B2$
3	CP063	$C \cdot B3 \cdot T4$
	CP030	$C \cdot B1 \cdot T4$
4	CP071	$C \cdot B3 \cdot T6$
	CP045	$C \cdot B2 \cdot T6$
	CP036	$C \cdot B1 \cdot T6$
	CP058	$C \cdot B4 \cdot T6$

enzyme elasticity components.

We found that the 76 control patterns of $C_{v_3}^{J_6}$ consist of 6 unique backbone patterns in combination with 13 different multiplier patterns as shown in Table 5.1. Multiplier patterns were categorised as either “T” multipliers or “B” multipliers, signifying that they consist of pathway components from either the top or bottom halves of the reaction scheme of the pathway shown in Fig. 5.2.

By identifying these backbone and multiplier patterns within the 11 dominant control patterns (Table 5.1) we found that the previously described control pattern groups align with the composition of the control patterns within a group. Patterns within the same group consist of the same backbone and T multipliers and only differ in terms of their B multipliers. Group 2 was the only group that did not contain a T multiplier, as its backbone pattern (B) extends into the same branch as the T multiplier patterns via acetolactate synthase. Apart from this exception, T multipliers were unique to each group, and groups 3 and 4 were the only ones to share the same backbone. The same B multiplier pattern, on the other hand, reoccurred multiple times between control patterns within different groups, with B3 forming part of the dominant control pattern in every group. Furthermore, only a subset of about half of the total multipliers and backbones ever appear in any of the important control patterns.

These results signify that control patterns that are dominant for the same conditions are also closely related in terms of their physical components. Intuitively this makes sense, as one would expect chains of local effects that follow mostly the same paths to behave similarly; however, overlap in terms of components is not necessarily a predictor of similar behaviour.

Single differences in composition between two patterns could result in major behavioural differences, while two unrelated patterns consisting of components with similar properties, such as being scaled by fluxes of similar magnitude or being similarly sensitive to their substrates or products, could also potentially behave similarly.

A ϕ_N parameter scan of the backbone and multiplier patterns further illuminates the source of variation between the different control patterns. Fig. 5.4A shows the values of the 6 backbones, each scaled by the common denominator, and clearly demonstrates the vast differences in terms of magnitude and response towards ϕ_N. Backbones C and F are a particularly illustrative example as they initially have relatively similar magnitudes and respond similarly towards an increase in ϕ_N; however, after $\phi_N \gtrsim 0.001$ their values diverge and by the end of the ϕ_N range they differ by ≈ 6 orders of magnitude. It is noteworthy that the similarity of C and F at low ϕ_N is purely coincidental, as C shares as many components with F as with any other backbone pattern.

The B and T multipliers are shown in Fig. 5.4B and C respectively. Within each of these groups, patterns respond significantly more homogeneously towards a change in ϕ_N than their backbone counterparts. This is particularly true for the B multipliers, which respond fairly uniformly towards increasing ϕ_N and differ chiefly in terms of magnitude. The T multipliers are more variable than the B multipliers, especially in terms of differences in magnitude; however, they exhibit a similar trend of decreasing magnitude in response towards increasing ϕ_N.

These results illustrate how the backbone and T multiplier patterns determine dominance of certain groups of control patterns over others while the B multipliers only determine the degree of dominance of different patterns within a group. Two explicit examples are shown in Fig. 5.5. In these figures, the use of logarithmic scales allows us to distinguish between the factors that act to increase the value of control pattern and those that act to decrease it. In Fig. 5.5A the control patterns 063 and 071, which have the largest magnitudes within groups 3 and 4, respectively, are compared. As visually depicted in Fig. 5.6, these patterns differ only in terms of their T multipliers, with T4 belonging to CP063 and T6 belonging to CP071. Comparing these T multipliers in Fig. 5.5A clearly shows how they are responsible for determining the values of CP063 and CP071. For the lower ϕ_N values where CP063 is dominant, the value of T4 is notably larger than T6, thus "pulling up" the value of this control pattern. However, as the ϕ_N increases, the magnitude of T4 decreases to below that of T6. In this case T4 has a greater negative effect on the value of CP063 compared to that of T6 on CP071. Thus, despite their predominantly similar compositions, CP063 and CP071 differ significantly in terms of their response towards ϕ_N.

In Fig. 5.5B a more complicated example is given. Here we compare CP071 with CP001,

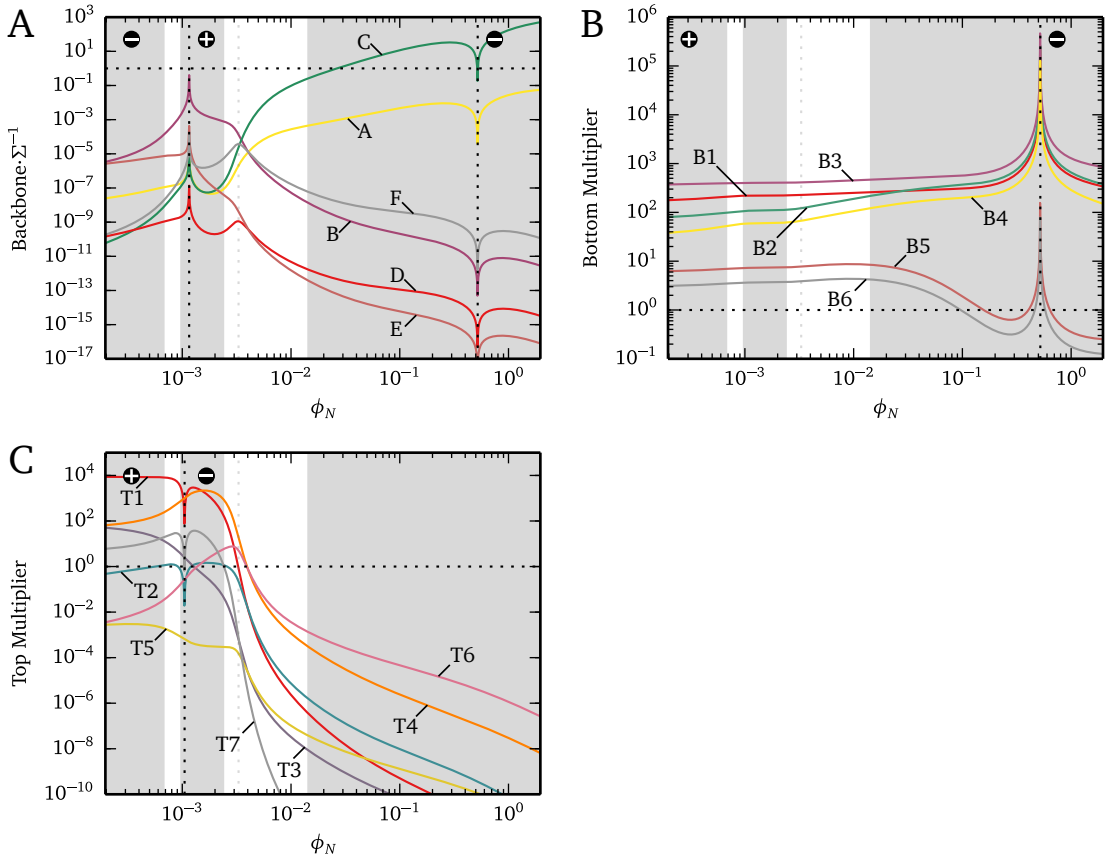


Figure 5.4: Backbone and multiplier patterns of the control patterns of $C^{J_6}_{\nu_3}$ as functions of ϕ_N. **A** The values of the 6 backbones (A–F) of $C^{J_6}_{\nu_3}$ scaled by the common denominator (Σ) of the control coefficients of this system. **B** The values of the 6 multipliers (B1–B6) that consist of components from the bottom half of the reaction scheme in Fig. 5.6. **C** The values of the 7 multipliers (T1– T7) that consist of components from the top half of the reaction scheme. Shaded regions indicate dominance by different control pattern groups as described in Fig. 5.3. Logarithmic coordinates are used together with a horizontal black dotted line to differentiate between multipliers and backbones that have an increasing (y-values > 1) or diminishing (y-values < 1) effect on their control pattern products. Absolute values of backbones, multipliers, and control patterns are taken to allow for plotting logarithmic coordinates, with crossover from positive to negative with increasing ϕ_N for multipliers and vice versa for control patterns (see Fig. 5.5) indicated by vertical dotted lines. Backbones each switch sign twice from a negative starting point on the left-hand side of **A**.

a control pattern with which it shares much less similarity than with CP063. In this case the patterns have both different backbones and different T multipliers; however, they once more share B3. A visual depiction of these control patterns is shown in Fig. 5.7. These patterns are dominant on opposite ends of the ϕ_N range (Fig. 5.3). Unlike CP063, CP001 is also much more comparable to CP071 terms of magnitude within the region that it dominates

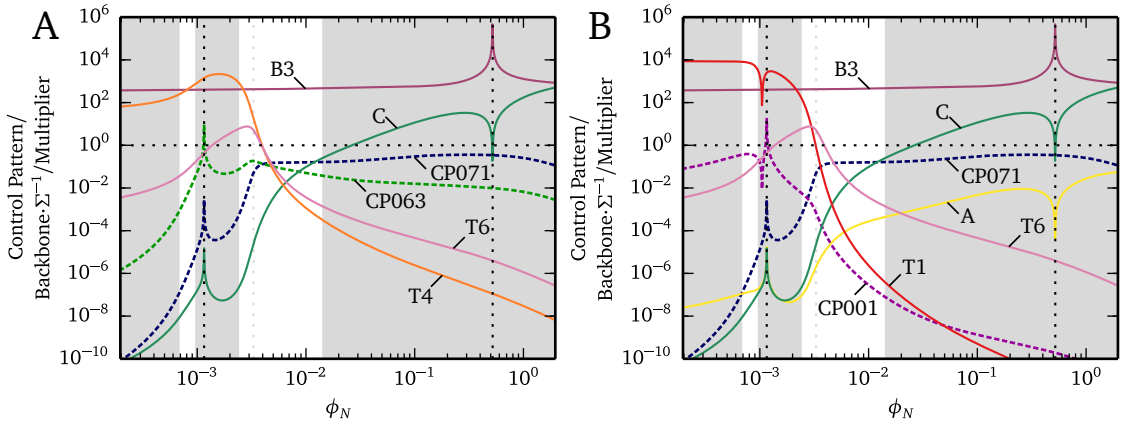


Figure 5.5: The backbone and multiplier components of the control patterns CP001, CP063, and CP071 of $C^{J_6}_{\nu_3}$ as functions of ϕ_N. **A** CP063 and CP073 together with their constituent backbone and multipliers. CP063 and CP071 differ only in their top half multipliers (T4 and T6). **B** CP001 and CP073 together with their constituent backbones and multipliers. CP001 and CP071 differ in terms of both their top half multipliers (T1 and T6), and their backbones (A and C). Backbone patterns are scaled by the common denominator (Σ) as in Fig 5.4. Shaded regions indicate dominance by different control pattern groups as described in Fig. 5.3. Logarithmic coordinates and absolute values of patterns are used in the same manner as in Fig 5.4 to indicate increasing or diminishing effects of backbone and multiplier patterns on the control patterns.

($\phi_N \lesssim 0.001$). From Fig. 5.5B it is clear that, while the value of the backbone A is significantly larger than that of C within this range, the dominance and magnitude of CP001 can largely be attributed to the value of T1, which is ≈ 3 orders of magnitude larger than T6. Additionally, patterns containing T1 are the only ones that serve to increase the value of CP001, save for the shared pattern B3. On the other hand, for high ϕ_N values both A and T1 have significantly lower values than their counterparts. For these ϕ_N values, T1 can once more be regarded as a dominant factor in determining the value of CP001, as its value is, in this case, ≈ 3 orders of magnitude smaller than any other pattern.

The concepts of multiplier and backbone patterns were immensely useful for indicating the source of the regime change mentioned earlier. While they were originally developed as stepping stones towards tracing full control patterns [121], as each multiplier and backbone pattern still represents an actual physical chain of effects within the pathway, they carry utility above their original intent and meaning beyond the mere numbers assigned to them. In the case of the T4 and T6 multipliers, which respectively determine CP063 and CP071, for instance, the differences in the values of these patterns comes from relatively minor differences in their compositions; T4 has a $J_{10}\varepsilon^{\nu_{10}}_{Acet}$ factor, whereas T6 has a $J_{11}\varepsilon^{\nu_{11}}_{Acet}$ factor (see Fig. 5.6). Thus, the dominance of either CP063 or CP071 depends on the magnitudes of two fluxes

and the sensitivities of two enzymes towards their substrates; as a change in ϕ_N affects the system, J_{10} and J_{11} are altered and the concentration of Acet changes, altering the values of $\varepsilon_{Acet}^{v_{11}}$ and $\varepsilon_{Acet}^{v_{10}}$. While, as noted, the situation is more complicated for the differences between CP001 and CP071, the same principle still holds.

Another interesting finding is that, while the properties of a backbone or multiplier pattern are determined by its composition, the metabolic branch in which the pattern resides also gives important information. We saw that both T and B multipliers respond similarly to changes in ϕ_N within their respective groups, in spite of some patterns having no shared components. One possible reason for this is that their properties are more similar than what is apparent from inspection due to certain flux relationships. In the case of the B multipliers, for instance, B1–B4 have much larger values than B5 and B6 for most of the tested ϕ_N values in spite of some of these patterns not sharing a single component. However, each of B1–B4 either contains a J_1 or a J_{12} factor, which are nearly equal for all tested ϕ_N values due to J_{12} being the only ATP demand flux within this pathway and J_1 being the greatest ATP supply flux (by orders of magnitude above J_5). Therefore, in practical terms each of B1–B4 share at least a single component. The two weakest patterns within the B group (B5 and B6) are thus those that contain a J_5 factor rather than a J_1 or J_{12} factor. For this same reason, B1–B4 exhibit a different response towards changes in ϕ_N than B5 and B6. In the T multiplier group a similar effect may also be at play.

In Section 5.3.5 we will explore these concepts further by examining the properties of the control patterns that bring about their behaviour.

5.3.3 Control patterns of $C_{v_6}^{J_6}$ and $C_{v_7}^{J_6}$

The other control coefficients that were important for determining the response of J_6 towards ϕ_N were $C_{v_6}^{J_6}$ and $C_{v_7}^{J_6}$. For these control coefficients 16 dominant patterns were identified using the same procedure as described in Section 5.3.1 for the $C_{v_3}^{J_6}$ patterns. The first notable finding was that each control pattern of $C_{v_6}^{J_6}$ corresponds with a pattern of $C_{v_7}^{J_6}$, differing only by a single elasticity coefficient. The $C_{v_6}^{J_6}$ control patterns can thus be transformed into their equivalent $C_{v_7}^{J_6}$ patterns by multiplying by the elasticity ratio $\varepsilon_{Acal}^{v_6}/\varepsilon_{Acal}^{v_7}$, and therefore we will only explicitly discuss $C_{v_6}^{J_6}$.

Interestingly, we found that curves representing the absolute percentage contribution for the $C_{v_6}^{J_6}$ control patterns over the ϕ_N range look very similar to those of $C_{v_3}^{J_6}$ as shown in Fig. 5.3A, and also exhibit the same 4 regions of dominance. Investigation of these dominant patterns revealed that each corresponds to a dominant pattern in $C_{v_3}^{J_6}$ and can be expressed in terms of these patterns in a similar manner to how $C_{v_7}^{J_6}$ can be expressed in terms of $C_{v_6}^{J_6}$. These control patterns are shown in Table. 5.3. In all but six cases, the backbone patterns of $C_{v_6}^{J_6}$

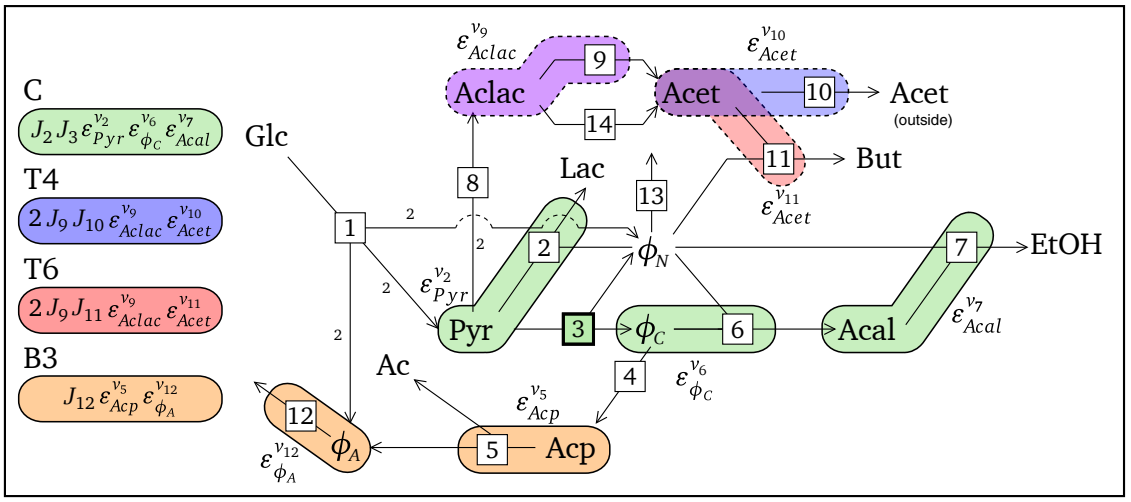


Figure 5.6: The components of control patterns 063 and 071 of $C^{J_6}_{v_3}$. The backbone and multipliers that constitute the dominant control patterns of group 3 and group 4 (CP063 and CP071) are indicated as groups of coloured bubbles that highlight their elasticity coefficient components as indicated by the key. These control patterns share the same backbone (C) and are therefore differentiated based on their respective incorporation of the multipliers T4 and T6 on the top half of the reaction scheme.

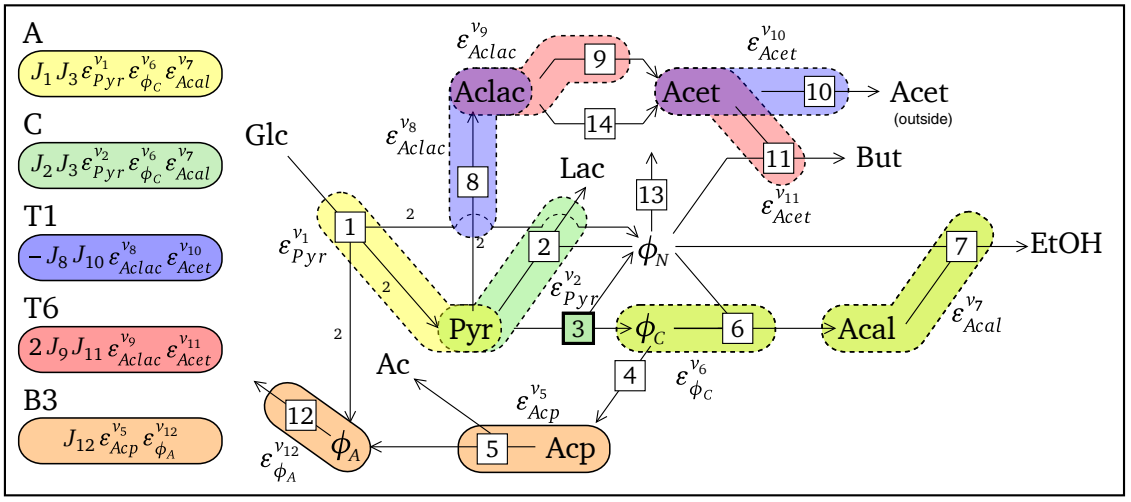


Figure 5.7: The components of control patterns 001 and 071 of $C^{J_6}_{v_3}$. The backbone and multipliers that constitute the dominant control patterns of group 1 and group 4 (CP001 and CP071) are indicated as groups of coloured bubbles that highlight their elasticity coefficient components as indicated by the key. These control patterns have different backbone patterns (A and C), as well as different multipliers (T1 and T6).

and $C^{J_6}_{v_3}$ differed by a single elasticity coefficient. In the six exceptional cases, the $C^{J_6}_{v_6}$ control patterns had a different backbone extending into the same reaction branch belonging to the

Table 5.3: Numerator expressions of the dominant control patterns of $C^{J_6}_{v_6}$ in terms of the $C^{J_6}_{v_3}$ control patterns. Each dominant control pattern of $C^{J_6}_{v_6}$ is, to some degree, equivalent to a dominant control pattern of $C^{J_6}_{v_3}$ and is expressed in terms of that pattern and the factors that account for the differences between the two expressions, considering only the numerators of the control patterns. These factors only affect the backbone patterns of the control patterns (see text). Control patterns are divided into groups based on their dominant ϕ_N value range, which corresponds with that of the dominant $C^{J_6}_{v_3}$ control patterns. Control patterns are arranged in descending order of relative importance within their group.

Group	$C^{J_6}_{v_6}$ control pattern	Expression
1	CP018	$\text{CP001} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP002	$\text{CP001} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
	CP001	$\text{CP011} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
2	CP068	$\text{CP064} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP087	$\text{CP031} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP105	$\text{CP064} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
	CP104	$\text{CP042} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP050	$\text{CP031} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
3	CP067	$\text{CP063} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP085	$\text{CP030} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP102	$\text{CP063} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
4	CP112	$\text{CP071} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
	CP060	$\text{CP036} \cdot (J_5\varepsilon^{v_4}_{\phi_C} / J_3\varepsilon^{v_6}_{\phi_C})$
	CP075	$\text{CP071} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP092	$\text{CP036} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$
	CP111	$\text{CP045} \cdot (-\varepsilon^{v_3}_{\phi_C} / \varepsilon^{v_6}_{\phi_C})$

B multiplier branch via phosphotransacetylase; however, the full control pattern expressions only differed by one scaling flux and one elasticity coefficient from those of $C^{J_6}_{v_3}$.

This result is striking, because it shows that while backbone patterns are unique to a control coefficient, the multiplier patterns can be shared among multiple control coefficients. The same multiplier patterns determined the dominance of certain control patterns over others for both $C^{J_6}_{v_6}$ and $C^{J_6}_{v_3}$. Clearly the actual values of $C^{J_6}_{v_6}$, $C^{J_6}_{v_7}$, and $C^{J_6}_{v_3}$ were not the same (see Fig. 4.8); however, their differences can be accounted for by rather minute differences

in their backbone patterns.

Before continuing our investigation of control patterns in Section 5.3.5 we will segue into the other important component of the partial response coefficients of $R^{J_6}_{\phi_N}$, namely the elasticity coefficients.

5.3.4 Explaining the behaviour of $\varepsilon^{\mathbf{v_6}}_{\phi_\mathbf{N}}$, $\varepsilon^{\mathbf{v_7}}_{\phi_\mathbf{N}}$, and $\varepsilon^{\mathbf{v_3}}_{\phi_\mathbf{N}}$

The elasticity coefficients $\varepsilon^{v_6}_{\phi_N}$, $\varepsilon^{v_7}_{\phi_N}$ and $\varepsilon^{v_3}_{\phi_N}$ represent the sensitivities of reactions through which changes in ϕ_N can be transmitted to the rest of the system (see Figs. 4.7 and 4.8). Our previous study clarified the role of these elasticity coefficient in eliciting the observed response of J_6 towards a change in ϕ_N; however, their respective behaviours remained unexplained [31] (Chapter 4). Here we will investigate these elasticity coefficients in terms of their distance from equilibrium and their binding and mass-action components.

One topic of interest in our previous study was the difference between the magnitudes of $\varepsilon^{v_6}_{\phi_N}$ and $\varepsilon^{v_7}_{\phi_N}$ [31] (Chapter 4). Fig. 5.8 shows that $\varepsilon^{v_6}_{\phi_N}$ is significantly larger than $\varepsilon^{v_7}_{\phi_N}$ for most of the ϕ_N range, which ultimately leads to a larger response of J_6 towards ϕ_N via reaction 6 than reaction 7 for most of the ϕ_N range, in spite of $C^{J_6}_{v_7}$ almost always being larger than $C^{J_6}_{v_6}$ (see Fig. 4.8). From the coloured regions demarcating where kinetic and thermodynamic control is expected, it is clear that reaction 7 (Fig. 5.8B) moves away from equilibrium much faster as ϕ_N increases than reaction 6 (Fig. 5.8A), becoming "far from equilibrium" [16] at a ϕ_N value nearly 2 orders of magnitude lower than the latter. This means that reaction 7 is expected to be kinetically controlled for most ϕ_N values, while reaction 6 is expected to be thermodynamically influenced for the same values.

The close to zero values of the binding elasticities $\varepsilon^{v_6\Theta}_{\phi_N}$ and $\varepsilon^{v_7\Theta}_{\phi_N}$ (Fig. 5.8A and B, purple lines), even far from equilibrium, seem to indicate that binding plays almost no role in determining the sensitivity of these reactions towards ϕ_N. Normally, as saturating concentrations of either a substrate or a product are approached, binding negatively affects the reaction rate as it approaches the limiting rate, leading to a negative binding elasticity coefficient (with a minimum value of -1) [16]. Here we see that as ϕ_N increases, the binding elasticity remains close to zero and barely changes, seemingly indicating that the enzymes never become saturated. In reality, however, this is an artefact from expressing the sensitivity towards both a substrate and a product in the form of a substrate/product ratio. In this case substrate and product binding counteract each other in each of reactions 6 and 7 due to these reactions each having similar affinities for both NADH and NAD^+. The slightly larger magnitude of $\varepsilon^{v_6\Theta}_{\phi_N}$ compared to $\varepsilon^{v_7\Theta}_{\phi_N}$ within their respective ranges of expected kinetic control is due to reaction 6 having a slightly higher affinity for NADH than NAD^+. It seems that for binding elasticity coefficients representing the sensitivities towards metabolite ratios to contribute

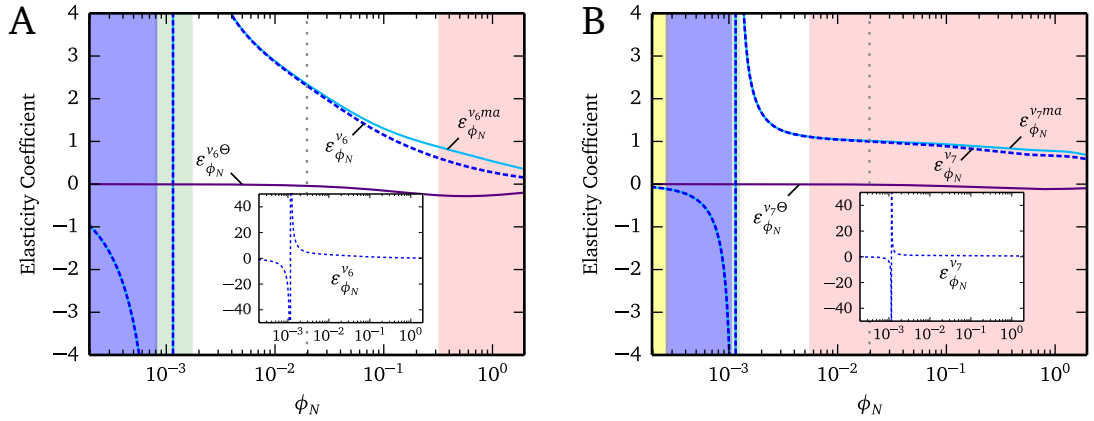


Figure 5.8: The elasticity coefficients $\varepsilon^{v_6}_{\phi_N}$ and $\varepsilon^{v_7}_{\phi_N}$ as functions of ϕ_N. **A** The elasticity coefficient $\varepsilon^{v_6}_{\phi_N}$ split into its binding and mass action components. The insert in **A** shows $\varepsilon^{v_6}_{\phi_N}$ on an expanded scale. **B** The elasticity coefficient $\varepsilon^{v_7}_{\phi_N}$ split into its binding and mass action components. The insert in **B** shows $\varepsilon^{v_7}_{\phi_N}$ on an expanded scale. The shaded areas in **A** and **B** indicate regions of expected kinetic and thermodynamic control of v_6 and v_7 (as per [16]), respectively with red : $\rho \leq 0.1$, white : $0.1 < \rho < 0.9$, green : $0.9 \leq \rho \leq 1/0.9$, blue : $1/0.9 < \rho < 1/0.1$ and yellow: $\rho \geq 1/0.1$.

significantly towards the total elasticity, their substrate and product affinities must differ by orders of magnitude. Assuming that the only way that the concentration of one member of a moiety-conserved cycle would increase on the metabolic time scale is if the concentration of its counterpart decreases, then it is reasonable to believe that in many real metabolic systems the moiety-conserved metabolites counteract the effect of binding of each other as demonstrated here.

As binding played a scarcely observable role, the difference in sensitivity of the two reactions largely comes down to the differences between $\varepsilon^{v_6 ma}_{\phi_N}$ and $\varepsilon^{v_7 ma}_{\phi_N}$ (Fig. 5.8A and B, light blue lines). These elasticities change sign at $\phi_N = 0.0012$ due to reactions 6 and 7 changing direction. This is reflected in the elasticity coefficients, which also change sign as the reactions switch from negative to positive. While both these elasticity coefficients move through the same values as ϕ_N increases, $\varepsilon^{v_7 ma}_{\phi_N}$ does so at a much faster rate relative to ϕ_N than $\varepsilon^{v_6 ma}_{\phi_N}$, which means that ultimately reaction 6 has a larger range of ϕ_N values for which it is very sensitive. This difference in the sensitivity ranges can be accounted for by the large K_{eq} of reaction 7 (1.24×10^4) compared to that of reaction 6 (1) [86], which resulted in a smaller ϕ value for the former for the same value of ϕ_N.

The behaviour of the final elasticity component of the three dominant partial response coefficients, $\varepsilon^{v_3}_{\phi_N}$, can be explained completely by the effect of binding, because this reaction is irreversible. This means that increasing the product ϕ_N has no effect on the reaction rate.

Like reactions 6 and 7, the effects of product and substrate binding counteract each other here; however, this reaction has an additional binding component in the form of inhibition by high ϕ_N values [86]. Like conventional binding elasticities, this inhibition elasticity varies between 0 and -1, and $1/K_i$ determines the value of ϕ_N where this elasticity has a value of -0.5.

From these results we can therefore conclude that the contributions of $\varepsilon^{v_6}_{\phi_N}$ and $\varepsilon^{v_7}_{\phi_N}$ towards the positive response of J_6 towards ϕ_N were largely a result of the mass-action effect, as binding played a negligible role. On the other hand, the fairly large negative value of $\varepsilon^{v_3}_{\phi_N}$ that increased together with increasing ϕ_N was purely due to binding as this reaction was modelled as being irreversible. In this way we could quantitatively clarify, at least partially, a high level property (namely flux-response towards increasing ϕ_N) in terms of the properties intrinsic to the reactions being catalysed, as well as the binding properties of the enzymes themselves. In the next section we will continue the search for origins of the control coefficients responsible for the J_6 flux response commenced in the previous sections.

5.3.5 Following the chains of effects

In this final section we will attempt to relate the properties of some the control patterns discussed earlier with those of their constituent elasticity coefficients. In this way we will show how the control coefficients are determined by the properties of the components of the system.

We will start with an example that relates to Section 5.3.3. As previously mentioned, control patterns of $C^{J_6}_{v_6}$ and $C^{J_6}_{v_7}$ differ by a single elasticity coefficient where each $C^{J_6}_{v_6}$ pattern numerator has a $\varepsilon^{v_7}_{Acal}$ factor, whereas those of $C^{J_6}_{v_7}$ have a $\varepsilon^{v_6}_{Acal}$ factor. Fig. 5.9 shows these two elasticity coefficients separated into their respective binding and mass-action components. We can clearly see the similarities in magnitude between $\varepsilon^{v_6}_{Acal}$ and $\varepsilon^{v_6}_{\phi_N}$ (Fig. 5.8A), and $\varepsilon^{v_7}_{Acal}$ and $\varepsilon^{v_7}_{\phi_N}$ (Fig. 5.8B), with $\varepsilon^{v_6}_{Acal}$ having a larger magnitude than $\varepsilon^{v_7}_{Acal}$ for $\phi_N \lesssim 0.03$ as in their counterparts. These similarities are to be expected, as we are dealing with the same reactions under the same conditions. Once more the mass-action effect plays the dominant role. However, unlike previously where substrate and product binding counteracted each other, the low values of $\varepsilon^{v_6\Theta}_{Acal}$ and $\varepsilon^{v_7\Theta}_{Acal}$ result from the concentration of Acal being far from saturating for both these reactions. Reaction 6 is even further from saturation as it has a over 300 times lower binding affinity for Acal. These results explain the previously observed differences between $C^{J_6}_{v_6}$ and $C^{J_6}_{v_7}$ in terms of these elasticity coefficients [31] (Chapter 4). The generally larger value of $C^{J_6}_{v_7}$ is due to $\varepsilon^{v_6}_{Acal}$ having a larger magnitude than $\varepsilon^{v_7}_{Acal}$. Similarly, while $C^{J_6}_{v_6}$ is smaller than $C^{J_6}_{v_7}$, it never reaches a value of zero because reaction 7 never becomes fully saturated by Acal.

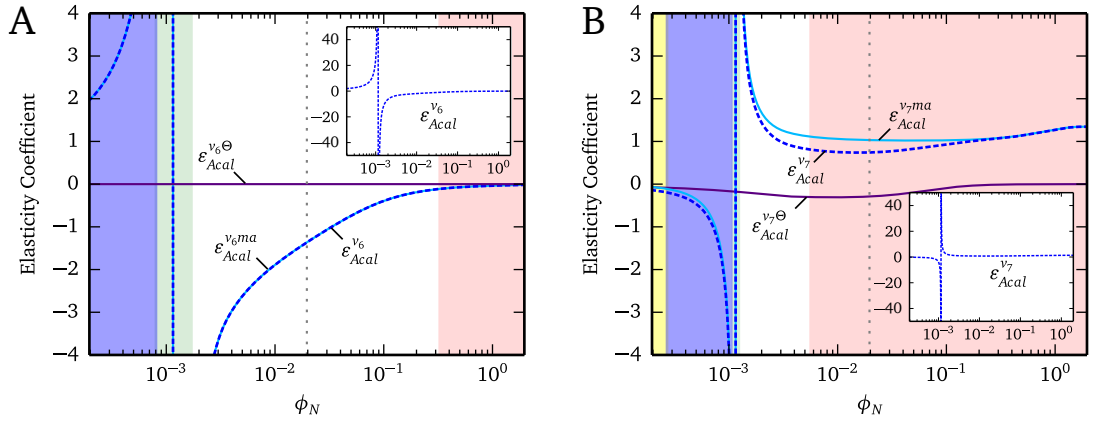


Figure 5.9: The elasticity coefficients $\varepsilon^{v_6}_{Acal}$ and $\varepsilon^{v_7}_{Acal}$ as functions of ϕ_N. **A** The elasticity coefficient $\varepsilon^{v_6}_{Acal}$ split into its binding and mass action components. The insert in **A** shows $\varepsilon^{v_6}_{Acal}$ on an expanded scale. **B** The elasticity coefficient $\varepsilon^{v_7}_{Acal}$ split into its binding and mass action components. The insert in **B** shows $\varepsilon^{v_7}_{Acal}$ on an expanded scale. The shaded areas in **A** and **B** indicate kinetic vs. thermodynamic control of v_6 and v_7, respectively with red : $\rho \leq 0.1$, white : $0.1 < \rho < 0.9$, green : $0.9 \leq \rho \leq 1/0.9$, blue : $1/0.9 < \rho < 1/0.1$ and yellow: $\rho \geq 1/0.1$. The vertical asymptote at 0.0012 indicates the change in direction of J_6 and J_7. This change in direction is also reflected in the elasticity coefficients.

This analysis clearly demonstrates the interplay between local and systemic properties. In the above example the properties of one reaction, to a large extent, determine the control properties of the other, and *vice versa*. While it is difficult to predict how the sensitivities of these reactions would be affected by changes in the moiety-conserved cycle represented by ϕ_N, as we have seen, some deductions can be made from the reaction properties regarding the magnitude and behaviour of the elasticity coefficients, which can potentially be translated into predictions regarding the control properties. For instance, we have seen that the lower magnitude of $\varepsilon^{v_7}_{Acal}$ compared to that of $\varepsilon^{v_6}_{Acal}$ for most ϕ_N values was a result of reaction 7 having a very large K_{eq} value. If this reaction were to be replaced with one that had a much lower K_{eq} value (all else being equal), $\varepsilon^{v_7}_{Acal}$ would potentially have a larger magnitude relative to $\varepsilon^{v_6}_{Acal}$ for the same ϕ_N values. Such a change would lower the value of $C^{J_6}_{v_6}$ by increasing the magnitude of the control coefficient common denominator, and increase the magnitude of $C^{J_6}_{v_7}$ by increasing the value of the value of the numerator in such a way that it partly counteracts the value of the denominator change.

Testing this hypothesis shows it to be true. For $\phi_N = 1.96 \times 10^{-4}$ the difference between $C^{J_6}_{v_7}$ and $C^{J_6}_{v_6}$ is the most pronounced in the reference state with values of 0.067 and 0.933, respectively. Replacing the kinetically controlled reaction 7 with a hypothetical reaction that has a thousand-fold smaller K_{eq} results in $\varepsilon^{v_7}_{Acal}$ having a larger magnitude relative to $\varepsilon^{v_6}_{Acal}$ for the reference state, which in turn leads to values of 0.403 and 0.445 for $C^{J_6}_{v_7}$ and $C^{J_6}_{v_6}$. While

this example is not very realistic as K_{eq} values are fixed, and the exact magnitude of such changes is difficult to predict by reasoning alone, it serves to illustrate that some rational predictions regarding the control properties of a system can be made by understanding how control coefficients are determined by the system structure and components, and how the relevant reaction parameters, based on the distance equilibrium, affect the behaviour of the reaction.

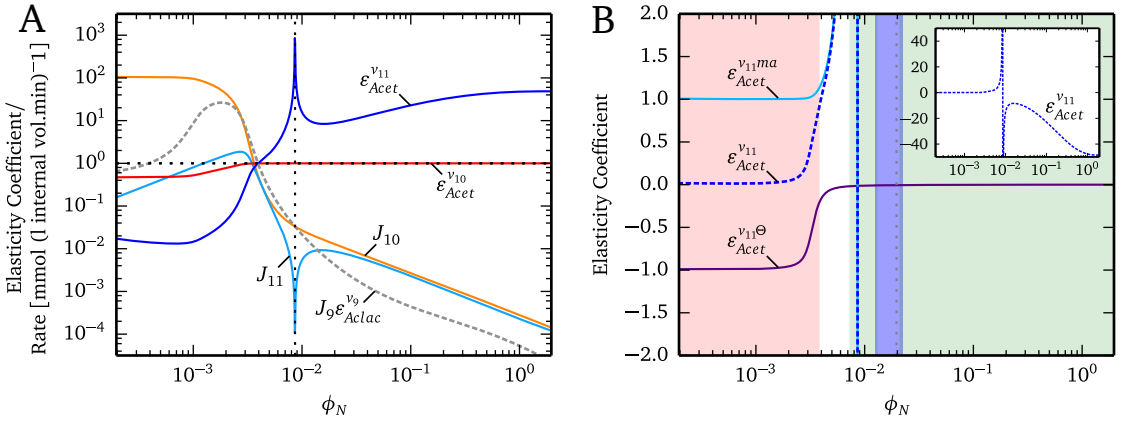


Figure 5.10: The flux and elasticity components of T4 and T6 as functions of ϕ_N. **A** shows the elasticity and flux factors that constitute the multipliers T4 and T6. Both multipliers share the factor $J_9\varepsilon^{v_9}_{Aclac}$, while T4 includes J_{10} and $\varepsilon^{v_{10}}_{Acet}$, and T6 includes J_{11} and $\varepsilon^{v_{11}}_{Acet}$. Logarithmic coordinates together with a horizontal black dotted line and absolute values are used in the same fashion as in Fig. 5.4 with the black vertical dotted line indicating the crossover from positive to negative values of J_{11} and $\varepsilon^{v_{11}}_{Acet}$. **B** The elasticity coefficient $\varepsilon^{v_{11}}_{Acet}$ split into its binding and mass action components. The insert in **B** shows $\varepsilon^{v_{11}}_{Acet}$ on an expanded scale. The shaded areas indicate kinetic vs. thermodynamic control of v_{11} with red: $\rho \leq 0.1$, white: $0.1 < \rho < 0.9$, green: $0.9 \leq \rho \leq 1/0.9$ and blue: $1/0.9 < \rho < 1/0.1$.

Our second example deals with the differences between control patterns of $C^{J_6}_{v_3}$ in groups 3 and 4, which were due to the differences between the multiplier patterns T4 and T6 (see CP063 and CP071 in Section 5.3.2) . As shown in Table 5.1, both pattern numerators consist of the factor $J_9\varepsilon^{v_9}_{Aclac}$ and T4 has the factor $J_{10}\varepsilon^{v_{10}}_{Acet}$, whereas T6 has $J_{11}\varepsilon^{v_{11}}_{Acet}$. In Fig. 5.10A, these different factors are shown individually against ϕ_N. At low ϕ_N values, both J_{10} and $\varepsilon^{v_{10}}_{Acet}$ have larger magnitudes than their counterparts in T6, which account for the dominance of patterns in group 3 over those in group 4 at these lower ϕ_N values. As ϕ_N increases, however, J_{10} and J_{11} become nearly equal in terms of value and response towards ϕ_N. This means that the observed regime change from group 3 to group 4 is determined by the difference in sensitivity of reactions 10 and 11 towards their substrate. Reaction 10 is irreversible and its substrate concentration is far from saturating at high ϕ_N, therefore the value of 1 of $\varepsilon^{v_{10}}_{Acet}$ is

exclusively a result of the mass-action effect. Reaction 11, on the other hand, is reversible, and as can be seen in Fig. 5.10B, changes from being far-from-equilibrium at low ϕ_N where Acet is saturating, to being close to equilibrium for $\phi_N \gtrsim 0.008$. This reaction remains close to equilibrium, because the concentration of its substrate Acet decreases to a larger degree than the co-subtrate ϕ_N increases, which leads to a large mass-action ratio.

Here we can see once more how the properties of single reactions can affect the system on a global scale. We can use our knowledge of the properties of these reactions to design a strategy for manipulating the control properties of the system. We can imagine that replacing the thermodynamically controlled reaction 11 with a reaction with a thousand fold larger K_{eq} should result in the new reaction becoming far-from-equilibrium for the same ϕ_N values, which will affect the values of J_{11} $\varepsilon_{Acet}^{\nu_{11}}$ and ultimately change the control properties of the system. In reality, however, the situation is not so simple. While a change in the K_{eq} of reaction 11 does cause the reaction to be far-from-equilibrium, which lowers the magnitude of $\varepsilon_{Acet}^{\nu_{11}}$ to be similar to $\varepsilon_{Acet}^{\nu_{10}}$, it also causes J_{10} to decrease in magnitude. These changes also affects the values of the other control patterns leading to all the control patterns having the same magnitudes as before the change! Similarly, completely knocking out reaction 11 causes the control patterns in group 3 (and other related control patterns) to completely compensate for the loss of function, ultimately resulting in no significant change in $C_{\nu_3}^{J_6}$. This illustrates that in spite of a rational approach, it remains difficult to predict the actual effect of a change.

As a third example we will revisit the control patterns 001 and 071 of Section 5.3.2. There are many more differences between these two patterns than there are between CP063 and CP071. In Fig. 5.11A the factors belonging to the backbone A and multiplier T1 are shown (CP001), whereas in Fig. 5.11B those belonging to backbone C and multiplier T6 are shown (CP071). The factors of B3 are not shown, as they are irrelevant for determining the differences between these two control patterns; however, they do overlap in terms of the factors $J_3\varepsilon_{\phi_C}^{\nu_6}\varepsilon_{Acal}^{\nu_7}$, which are collected into a single term here. Focussing on the left-hand side of Figs 5.11A and B where CP001 dominates, it is clear that this control pattern has more additive factors than its counterpart (J_1, J_8, and J_{10} vs. J_2, J_9). Additionally, for these low ϕ_N values, the subtractive factors of CP001 have magnitudes closer to one than those of CP0071, and thus do not affect CP001 as severely as the factors of CP0071. As ϕ_N increases, however, the situation reverses completely. On the right-hand side of Figs 5.11A and B where CP0071 is the dominant pattern, the once positive factors of CP001 have decreased by more than 6 orders of magnitude, leaving this control pattern with a single additive factor (J_1). CP071, on the other hand, only has two negative factors (apart from the shared factor) while all other factors are either equal to one or larger than one, and thus additive.

Unlike the previous examples, the differences between these factors are much more con-

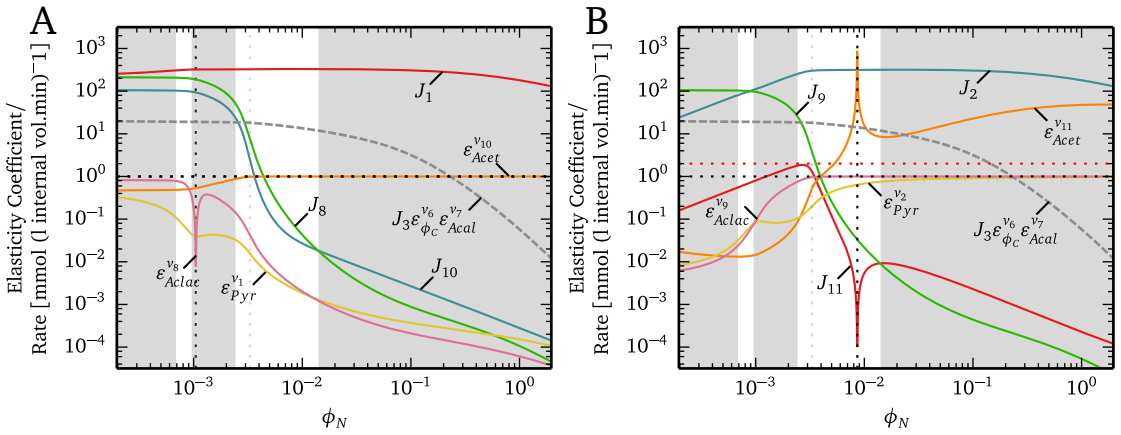


Figure 5.11: The flux and elasticity components of T1×A and T6×C of $C^{J_6}_{v_3}$ as functions of ϕ_N. The elasticity and flux factors that constitute the multipliers T1 and T6, as well as the backbones A and C are shown. **A** shows elasticities and fluxes of T1 and A, whereas **B** shows those of T6 and C. Both T1×A and T6×C share the factor $J_3\varepsilon^{v_6}_{\phi_C}\varepsilon^{v_7}_{Acel}$. Logarithmic coordinates together with a horizontal black dotted line and absolute values are used in the same fashion as in Fig. 5.4. Shaded regions indicate dominance by different control pattern groups as described in Fig. 5.3. The black vertical dotted line in **A** indicates the crossover from negative to positive values of $\varepsilon^{v_8}_{Aclac}$ (due to product activation as a result of cooperative product binding in the reversible Hill equation [80]) and **B** indicating the crossover from positive to negative values of J_{11} and $\varepsilon^{v_{11}}_{Acet}$.

voluted and difficult to explain. Firstly, the decreases in J_8, J_9, and J_{10} after their initial large values are all related to the increase in J_2 as ϕ_N increases, which shifts flux away from these reactions. While J_{11} was never large to begin with, its decrease is also related to the shift in flux away from the metabolic branch J_2, and its elasticity $\varepsilon^{v_{11}}_{Acet}$ increases due to the reaction being near-equilibrium as previously discussed. These changes in flux also have an effect on the elasticity coefficients. The elasticity $\varepsilon^{v_8}_{Aclac}$ decreases in magnitude as ϕ_N increases due to the decreasing concentration of Aclac, which is in turn a result of the decreased J_8 flux. The decrease in Aclac also causes $\varepsilon^{v_9}_{Aclac}$ to increase up to a value of one. Due to the fact that this reaction is irreversible, $\varepsilon^{v_9}_{Aclac}$ never exceeds one, even in the face of the extremely low substrate concentration. A similar situation occurs for $\varepsilon^{v_1}_{Pyr}$ and $\varepsilon^{v_2}_{Pyr}$, where the decrease in Pyr causes the decrease in the former, whereas the latter increases up to one due to its irreversibility. Thus ultimately, the broad shift in dominance from CP001 (and CP011) to CP071 (and other patterns in group 4), can be related back to the shift in flux from J_8 to J_2.

These results illustrate the relationship between the high-level properties of the system, and those of the components. We have quantified the previously observed flux response in terms of the chains of effects that constitute its control coefficient components. In this way, we go beyond what is offered by numeric control analysis by providing a much more

complete picture of the observed effect, thus leading to a significantly deeper understanding of the machinery responsible for the system behaviour. To our knowledge, this is the first application of symbolic control analysis [23] and control pattern analysis [17, 119] in which control patterns were investigated in this manner.

As mentioned in the Introduction to this chapter, the exact cause of the shift from homolactic to mixed-acid fermentation in pyruvate branch metabolism remains an open question [224]. While, as previously noted [31] (Chapter 4), this model does not yield all of the results expected from a shift in the redox-balance of this system, the shift away from mixed-acid fermentation at high NADH/NAD$^+$values is observed [210–213]. In our previous work we focussed specifically on the flux towards ethanol and acetaldehyde and its decrease at high NADH/NAD$^+$values. Here we have shown that as NADH/NAD$^+$increases from very low values, there is also a shift from flux towards acetoin and butanediol towards lactate. What is particularly interesting here is how these shifts are related to one-another. The acetoin/butatediol shift towards lactate altered the control properties of the system in such a way that promoted the originally observed decrease in ethanol and acetaldehyde in response to an increase in ϕ_N by enabling the high value of $C^{J_6}_{v_3}$ (and thus the high value of ${}^{v_3}R^{J_6}_{\phi_N}$). While this was not the only effect in play, it is striking to be able to relate such events with one another using a quantitative basis. We do not, however, believe that this is a complete picture of the shift in from homolactic to mixed-acid fermentation, as this model is relatively limited in terms of its scope and omits certain details such as the inhibition of lactate dehydrogenase, alcohol dehydrogenase and glyceraldehyde 3-phosphate dehydrogenase by ATP and ADP [208, 209]. In addition it does not fully include glycolysis, which would significantly increase the number of chains of local effects and thus allow for even more complex behaviour. This work rather provides a possible entry point into illuminating this phenomenon.

5.3.6 Control patterns in the free-ϕ_N model

One final question that remains is the issue of the use of the fixed-ϕ_N vs. the free-ϕ_N model. Similar to Table 5.3, which shows the control patterns of $C^{J_6}_{v_6}$ in terms of those of $C^{J_6}_{v_3}$, Table 5.4 shows the $C^{J_6}_{v_3}$ patterns in the free-ϕ_N model in terms of $C^{J_6}_{v_3}$ in the fixed-ϕ_N model. While there are differences between the dominant control pattern expressions, these are relatively minor, and in most cases restricted to two symbols. More convincing perhaps, is the similarity of the contributions of these individual patterns towards the total control as shown in Appendix C. In this appendix we the show fluxes and the control pattern contributions towards $C^{J_6}_{v_3}$ (similar to Fig. 5.3) in the free-ϕ_N model as a function of ϕ_N. Ultimately, while fixing ϕ_N does alter the structure and control properties of the pathway, in this specific case these alterations were minor and did not impact upon our main conclusions. However, this

may not be the case for other models.

5.4 Conclusion

In this chapter we have revisited results generated in a previous study in order to provide a demonstration of the analysis of the regulation of a system in terms of the properties of its components. Previously, we have shown that the negative flux-response of the acetaldehyde dehydrogenase flux due to an increase in the ratio of NADH to NAD^+ was a result of interactions of these intermediates with other reactions within a model of *L. lactis* pyruvate branch metabolism. We subsequently explained the influence of these reactions on the acetaldehyde dehydrogenase flux in terms of their sensitivities towards this metabolite ratio, and in terms of the control of these reactions over the acetaldehyde dehydrogenase flux. Here, we dissected these reaction sensitivities and flux-control coefficients in order to arrive at a mechanistic explanation of the previously described flux response. We have also derived some general conclusions regarding the use of control patterns in the investigation of metabolic systems.

Our analysis of control patterns of this system was greatly simplified by using the concepts of backbone and multiplier patterns, which provided a descriptive language for comparing and contrasting control patterns that is much more digestible than considering the full control pattern expressions. We have seen that certain combinations of interactions between components in a system are more important than others for determining the control of a reaction over a system variable, thus the number of control patterns to consider is reduced to only those that actually contribute to the control. The importance of these control patterns is not static, and shifts as parameters are changed; however, certain control patterns never contribute significantly towards the control regardless of a specific parameter value. We also found that different reactions have similar control patterns for the same metabolic variables.

In addition to describing the control coefficients in terms of chains of physical effects, we could, in a select few cases, explain certain aspects of their behaviour in terms of the properties and behaviour of single reactions. The information gleaned from describing control in terms of a single elasticity coefficient can probably not be used reliably to alter the control properties of a system, as we have seen. However, by zooming out and considering the general trends that groups of elasticity coefficients indicate (in the form of control patterns), these techniques could aid in metabolic engineering by guiding towards rational engineering strategies.

Many of the general findings described here are certainly not limited to the model we have used, as they are based on the well established principles of metabolic control analysis, simply seen from a different perspective. This work was made possible by software implementations

Table 5.4: Numerator expressions of the dominant control patterns of $C^{J_6}_{\nu_3}$ in the free-ϕ_N model in terms of the $C^{J_6}_{\nu_3}$ control patterns in the fixed-ϕ_N model. Each dominant control pattern of the free-ϕ_N model is, to some degree, equivalent to a dominant control pattern in the fixed-ϕ_N model and is expressed in terms of that pattern and the factors which account for the differences between the two expressions, considering only the numerators of the control patterns. Control patterns are divided into groups based on their dominant ϕ_N value range, which correspond with those of the dominant control patterns in the fixed-ϕ_N model. Group 2 is divided into smaller groups A and B, as there is a difference between the shape of the patterns in these groups (See Fig. C.4). Control patterns are arranged in descending order of relative importance within their group.

Group	Control Pattern	Expression
1	CP003	$\text{CP001} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP027	$\text{CP011} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
2A	CP177	$\text{CP064} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP088	$\text{CP031} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP116	$\text{CP042} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
2B	CP181	$\text{CP064} \cdot \left(J_{2}\varepsilon_{\phi_N}^{\nu_{2}}\right)$
	CP092	$\text{CP031} \cdot \left(J_{2}\varepsilon_{\phi_N}^{\nu_{2}}\right)$
	CP120	$\text{CP042} \cdot \left(J_{2}\varepsilon_{\phi_N}^{\nu_{2}}\right)$
3	CP176	$\text{CP063} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP214	$\text{CP063} \cdot \left(2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}/\varepsilon_{\phi_C}^{\nu_6}\right)$
	CP069	$\text{CP063} \cdot \left(-2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}/(\varepsilon_{\phi_C}^{\nu_6}\varepsilon_{Acal}^{\nu_7})\right)$
	CP087	$\text{CP030} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
4	CP216	$\text{CP071} \cdot \left(2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}/\varepsilon_{\phi_C}^{\nu_6}\right)$
	CP067	$\text{CP036} \cdot \left(2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}/\varepsilon_{\phi_C}^{\nu_6}\right)$
	CP071	$\text{CP071} \cdot \left(-2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}/(\varepsilon_{\phi_C}^{\nu_6}\varepsilon_{Acal}^{\nu_7})\right)$
	CP212	$\text{CP058} \cdot \left(2J_4\varepsilon_{\phi_N}^{\nu_1}\varepsilon_{\phi_A}^{\nu_5}/\varepsilon_{\phi_A}^{\nu_1}\right)$
	CP191	$\text{CP071} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP211	$\text{CP036} \cdot \left(-2J_4\varepsilon_{\phi_C}^{\nu_4}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}/(\varepsilon_{\phi_C}^{\nu_6}\varepsilon_{Acal}^{\nu_7})\right)$
	CP098	$\text{CP036} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$
	CP124	$\text{CP045} \cdot \left(J_{13}\varepsilon_{\phi_N}^{\nu_{13}}\right)$

of theoretical tools that were not available 10 years ago and we believe that the techniques described here can lead to a richer understanding of why and how other systems behave the way that they do. In our case, the application of these techniques not only led to biological insight, but also to new insights regarding the further development of our analysis software (as will be discussed in Section 6.3).

Chapter 6

General Discussion

This final chapter offers a broad overview and consolidated discussion of the work presented thus far. In Section 6.1 a summary of the main results of Chapters 3–5 is presented, whereafter this work is critically appraised and discussed in terms of its general implications within the context of systems biology and metabolic pathway analysis in Section 6.2. In Section 6.3 suggestions are made for possible future developments, before finally concluding in Section 6.4.

6.1 Synopsis

There were two interrelated primary motivations for the work presented in this dissertation. Firstly, we wanted to develop a software "toolbox" of implementations of various previously conceived metabolic analysis frameworks. Each of these tools had to be individually useful for exposing previously unknown properties of metabolic models relating to their behaviour, control, and regulation; however, a key envisioned strength of these tools was their potential for being applied to complement each other in this undertaking. Secondly, as we perceived most of the analysis frameworks included in our software toolbox as being underutilised in the investigation of models of real metabolic systems within the literature, we wanted to explore the practical utility of our tools in this domain. Thus, we set out to utilise these tools in the analysis of previously published kinetic models in order better understand the systems they represent.

After a review of the relevant literature, we presented our software package `PySCeSToolbox` in Chapter 3. This software includes three main tools, with each being a software implementation of a conceptual framework that can be utilised for model analysis. These three tools are `RateChar`, `SymCa`, and `ThermoKin`. `RateChar` automatically generates rate-characteristic plots of the supply and demand blocks of the variable intermediates of metabolic models in

order to perform generalised supply-demand analysis [15]. `SymCa` is used to determine the symbolic expressions of the control coefficients of a metabolic system in terms of the elasticity coefficients of its steps; therefore this tool provides the functionality to perform symbolic metabolic control analysis [23, 24] and control pattern analysis [17, 119, 121]. Finally, `ThermoKin` can be used to determine the contributions of the thermodynamic and kinetic aspects of rate equations towards the rates of the reactions they describe in a metabolic pathway [16]. Additionally it can determine the degree to which these terms control the reaction rate, as well as the distance of the reaction from equilibrium. While each of the software tools provided by `PySCeSToolbox` can be used in isolation, they are designed in a way that facilitates their combined use in metabolic model analysis.

In Chapter 4, we utilised `RateChar` to perform generalised supply-demand analysis on two previously published metabolic models. One of these models was of pyruvate branch metabolism in *Lactococcus lactis* [86], whereas the other was of aspartate-derived amino-acid synthesis in *Arabidopsis thaliana* [85]. Here we focussed specifically on elucidating the regulatory effects of multiple routes of interaction between various reaction blocks in each model. In the pyruvate branch model we explored the regulatory effects of the moiety-conserved cycles of ATP/ADP and NADH/NAD^+. One notable result was that a demand block of NADH/NAD^+ (with the demand enzyme acetyldehyde dehydrogenase) exhibited a response reminiscent of substrate inhibition towards in increase in the ratio NADH to NAD^+. This effect was, however, shown rather to be a result of the interaction of the members of this moiety-conserved cycle with pyruvate dehydrogenase upstream from acetyldehyde dehydrogenase.

In the aspartate-derived amino acid synthesis model, we concentrated on the regulation of the aspartate semialdehyde supply block by this intermediate itself. This supply block was found to respond negatively towards a positive perturbation of aspartate semialdehyde due to the interaction of this intermediate with its demand blocks, rather than through the expected product inhibition of the supply enzyme. This phenomenon was a result of an increase in the concentration of intermediates within the aspartate semialdehyde demand blocks, which act as upstream inhibitors of the supply block. Thus, for both models investigated within this chapter, the presence of multiple routes of interaction between reaction blocks resulted in unexpected regulatory phenomena that were uncovered with, and subsequently quantified using, generalised supply-demand analysis.

The analysis involving the pyruvate branch model was extended in Chapter 5. Here we first used `SymCa` to explore the control properties of this system, focussing specifically on the flux control of the acetyldehyde dehydrogenase NADH/NAD^+ demand block. By focussing on the steps that dominated the overall flux control of this block, our analysis revealed that, out of the large variety of possible chains of local effects that can propagate from these steps,

only a select few control patterns were actually responsible for the observed flux control. Furthermore, there was a degree of overlap between the chains of local effects of the dominant control patterns, i.e., the most important paths for propagating perturbations in terms of their contributions towards the observed control coefficients were closely related, and they could therefore be categorised into groups of related patterns. Secondly, we utilised `ThermoKin` to investigate the thermodynamic and kinetic aspects of the constituent elasticity coefficients of a selection of the aforementioned dominant control patterns. In these modelling experiments we demonstrated how the differences in some control coefficients can be traced to the properties of specific individual enzyme-catalysed reactions. By making appropriate alterations to these reactions according to their thermodynamic and kinetic properties, we were able to alter the flux control in one case, but not in another. Ultimately, the analyses in this chapter revealed the underlying mechanistic basis of the observed flux responses of the previous chapter.

6.2 Significance and critique

Over the last two decades, purely structural metabolic models have become increasingly prevalent. However, a survey of the curated models hosted by the Biomodels database [159] reveals that since 2010 at least 200 kinetic models have been published. While this number includes models other than metabolic models (e.g. [225, 226]), it also includes a variety of relatively large metabolic models (e.g. [227–229]). Regardless of the size of the model, however, after publication many of these models are not utilised in further investigations. This represents a significant waste of resources as, in our opinion, model construction is only the first step of the "modelling approach". As stated in the introduction of this text (Chapter 1), a fundamental goal of systems biology is to reach a system-level understanding of biology by concerning itself with how the properties of, and non-linear interactions between, the molecular components of biological systems give rise to higher levels of function. Throughout the remainder of this text we have endeavoured to reach system-level understanding by developing software implementations of metabolic analysis frameworks, by applying these tools in a novel manner, and by uncovering previously unknown behaviour of two metabolic models.

While the methods contained within `PySCeSToolbox` have been utilised in the past, and it is technically possible in principle to apply them in model analysis without using `PySCeSToolbox`, the development of specific software implementations of the thermodynamic/kinetic analysis framework, symbolic control analysis, and generalised supply-demand analysis have unlocked the potential of these frameworks in a way that was previously inaccessible. One notable strength of these software implementations is that they are completely

general. This means that, while there are some limitations in terms of the size of the models that can comfortably be approached (see discussion below), each of the tools provided by `PySCeSToolbox` can be utilised to investigate metabolic models of arbitrary size and complexity using essentially the same procedure. Secondly, all three of `RateChar`, `SymCa`, and `ThermoKin` are to a large extent automated, which means that minimal user intervention is required when utilising these tools. These two features abolish the need to develop specialised solutions for individual models, thus greatly reducing the time-investment and technical barrier for performing the analyses provided by `PySCeSToolbox`.

A second advantage stems from the fact that `PySCeSToolbox` is implemented as a Python library. While this decision may be a point of contention among potential users of the software, as it necessitates a scripting approach to model analysis, we believe that it is one of the most powerful assets of this software. Together with the fact that the tools of `PySCeSToolbox` are general and automated, a key benefit to this approach is the high degree of reproducibility and repeatability that it imparts to modelling experiments; by removing all need for user intervention after a script is developed, *in silico* experiments and analyses may be executed multiple times on the same computer, or on different computers, with no variation in output. Thus, assuming the original analysis script is correct, the produced results will always be accurate. In contrast, these highly sought-after traits of reproducibility and repeatability are much more difficult to achieve in simulators and tools that require user guided analysis through continuous interaction with a graphical user interface or a command-line interface, as this process is prone to user-error.

In Chapters 4 and 5, the features described in the above two paragraphs were indispensable to our own analyses. In Chapter 4 the scripting approach allowed us to perform practically the same analysis on models of completely different systems (pyruvate branch metabolism vs. aspartate-derived amino-acid synthesis) using the largely the same scripts with only minor alterations. Moreover, in the case where different versions of the same model were investigated (such as the various knockout models of Chapter 4, or the fixed moiety ratio and free moiety ratio models of Chapter 5), alterations to scripts were even more sparse. This greatly simplified the execution of modelling experiments, thus freeing up mental energy for interpreting results, rather than generating them.

Each of the metabolic analysis frameworks implemented in `PySCeSToolbox` has distinct uses for uncovering various properties of metabolic systems that might not be accessible using other techniques. GSDA, for instance, was originally developed in order to give researchers an entry point into the investigation of metabolic models. It achieves this by giving a broad overview of the behaviour, control, and regulation of a system, and by highlighting its most important regulatory features. Thus, while this framework relies on the principles of MCA,

it offers a more comprehensive view than what is achievable through investigation of the control and elasticity coefficients of a system. In Chapter 4, we have specifically used it to illuminate the regulatory role of multiple routes of interaction between reaction blocks in two distinct metabolic models. While it would have been possible to generate these results purely through the application of MCA, this process would also have been much more laborious as it would have involved using different versions of the models (one for each fixed variable intermediate) and a large number of calculations.

Similarly, symbolic control analysis offers advantages over numeric control coefficient analysis. In addition to indicating the degree of control that reaction steps have over the metabolic variables of a pathway, it illuminates the source of this control. Thus symbolic control analysis has far greater explanatory power than "conventional" MCA. In Chapter 5, this framework was central to uncovering the chains of local effects that resulted in the negative response of the acetyldehyde dehydrogenase NADH/NAD^+ demand block towards an increase in its substrate presented in Chapter 4. Here the control coefficient expressions themselves provided insight into a variety of phenomena. Firstly, they indicated that different chains of local effects determined the control properties of the system under different conditions. Secondly, by investigating and comparing the properties of the components of these chains, they allowed us to understand why different chains of local effects were dominant under different conditions. Finally, performing these same comparisons between the chains of local effects of different control coefficients pointed out that the source of their different values were related to the composition of their control patterns (and subsequently the properties of the components of these control patterns). In summary, the symbolic control coefficient expressions are able to quantify the control of a system in terms of its components and composition.

While software providing functionality for performing symbolic control analysis has existed for more than 20 years (albeit in a much more limited capacity than presented here), very few studies have, to our knowledge, utilised these tools. In the few cases where they have been used, the symbolic control coefficient expressions were never used in an explanatory capacity to the extent presented here. For instance, in a previous study by Thomas *et al.* [27], `MetaCon` [22] (see Section 2.4) was used to generate expressions for the control coefficients of glycolysis in *Solanum tuberosum*. These authors subsequently substituted the values of the elasticity coefficient of the steps of the system into the control coefficient expressions to determine their values. By varying the value of a particular elasticity coefficient they could determine its effect on important control coefficient values. While keeping in mind that these results were generated nearly two decades ago, it is noteworthy that the same results could have been achieved through inversion of different numeric **E**-matrices for each set of elas-

ticity coefficients [57]. Thus, in this case the control coefficient expressions were not really essential to the study. This is in stark contrast to the manner in which control coefficient expressions were utilised in Chapter 5 and in the case studies performed by Akhurst [24] in his thesis on the development of the original incarnation of `SymCa`. In both of these cases, the control coefficient expressions and their control patterns were used to relate higher level system behaviour to the chains of interactions between the physical components of the system. We therefore believe that this framework has been severely underutilised in the past.

While the thermodynamic/kinetic analysis framework of Rohwer and Hofmeyr [16] offers an arguably more justified demarcation between far-from and near-equilibrium reactions than that suggested by Rolleston [135], it also provides a more rigorous approach to investigating the effects of intrinsic mass-action and enzyme binding and catalysis. In contrast to approaches that rely only on the "distance-from-equilibrium" metric to classify a reaction rate as being determined by mass-action or by enzyme kinetics, this framework utilises elasticity coefficients to quantify the specific effects of these components. In our investigation presented in Chapter 5, we used this technique to explain the differences observed in the elasticity coefficient components of the partial response coefficients of the acetyldehyde dehydrogenase $NADH/NAD^+$ demand block towards $NADH/NAD^+$ of Chapter 4. In this case, however, the kinetic contribution towards determining the reaction sensitivity was zero, due to the fact that the elasticity coefficients were determined with respect the ratio of substrate to product rather than for an individual intermediate.

As previously mentioned, a particular strength of above-mentioned tools is their potential for complementing one another. In this dissertation this is exemplified in our analysis of the pyruvate branch metabolism model presented in Chapters 4 and 5. GSDA was used as an entry point in determining the behaviour of this system, and subsequently followed by analyses using `SymCa` and `ThermoKin`, which revealed finer details. Ultimately, this strategy uncovered high-level system behaviour and explained this behaviour in terms of both the interactions between system reactions, and the individual properties of the reactions. We believe that this fulfils the criteria for a mechanistic explanation of a biological system where neither the individual components, nor the complete system itself, is privileged [230].

This is not to say that these techniques are without limitations or caveats. In GSDA the intermediates are fixed and varied over a range to generate rate characteristic plots. This process has two important implications. Firstly, it prevents communication between reaction blocks through this intermediate, which results in different control properties for the "fixed" vs. "unfixed" models. Secondly, this artificial perturbation cannot be counteracted by the system as would be expected in regular dynamically stable systems. In our analysis in Chapter 4 we utilised two fixed intermediate models as part of a GSDA to investigate the flux-

responses of different reaction blocks towards the fixed intermediates, and to subsequently explain these flux-responses in terms of the different routes of interaction between the fixed intermediates and the reaction blocks. In this case, parametrisation of the intermediates was necessary to isolate the effect of changes in their concentrations on the system without the system counteracting these changes. Similarly, in Chapter 5 the chains of local effects that comprise the control coefficient components of the flux-responses investigated in the previous chapter were analysed using a model in which the NADH/NAD^+ ratio was fixed. As described in Section 5.3.6, we justified our use of the fixed model in two ways. Firstly, we wanted to extend the analysis of the previous chapter, thus warranting the use of the same model. Secondly, in spite of altering the control properties of the system, very similar behaviour and dominant control patterns were observed in the reference model when modulating of the activity of NADH oxidase. However, it is important not to overextend this approach and to be mindful of the consequences of the technique; in other models, parametrising an intermediate of the system may lead to vastly different behaviour and control properties compared to the reference model, thus limiting the usefulness of the "fixed" model for explaining actual biological phenomena.

A possible issue with symbolic control analysis is the practicability of the task of sifting through and making sense of the generated control patterns that constitute the control coefficients of a system. For instance, while the model of pyruvate branch metabolism consists of only 14 reactions, more than 200 control patterns were generated in the reference model for $C^{J_6}_{v_6}$ (Chapter 5), and one can therefore expect an enormous number of patterns for larger models. Our solution was to utilise backbone and multiplier patterns to simplify the task of analysing and comparing control patterns. Moreover, by focusing on the model where the NADH/NAD^+ ratio was fixed, the number of control patterns was significantly reduced. Nevertheless, this technique might not be viable in all cases due to the previously described issues. Finally, all control patterns containing elasticity coefficients of zero value (indicating insensitivity of a reaction towards a substrate or product) were disregarded, which further decreased the number of control patterns. This strategy can be extended to disregarding control patterns that contain elasticity coefficients known to be very close to zero under the conditions pertinent to a specific investigation. Another viable option may be to consider groups of reactions as single units using group elasticities as originally suggested by Hofmeyr [121]. In this strategy multiple reactions are effectively regarded as a single reaction with a single elasticity coefficient. Utilising the concept of monofunctional units, which are groups (or blocks) of reactions that only interact with the rest of the system through a single degree of freedom [231], can also be useful as they will most probably result in the simplest control coefficient expressions. Thus, analysis of a model could proceed in stages where the full

complexity is abstracted until there is a specific need for investigating the specific chains of local effects contained within the block of reactions represented by a group elasticity.

Finally, one should also be mindful of the fact that the results produced by these techniques are to a large extent only a reflection of the properties of the system under the specific configuration and conditions when they were generated. In other words, while they explain why a system behaves as it does, they cannot always reliably be used to predict the outcome of an alteration to the system. An example of such a case can be found in our investigation of the differences between control patterns 063 and 071 of $C^{J_6}_{\nu_3}$ in the pyruvate branch model in Section 5.3.5. Here, replacing one reaction with another with the goal of modifying the control properties of the system was unsuccessful, as the system-level effect of our alteration was ultimately counteracted by the system. Thus, while we could successfully explain the behaviour of the system in terms of its component properties and interactions, manipulations based on this information did not result in system behaviour matching our expectations. One must therefore avoid the temptation to fall into the reductionist trap of traditional molecular biology of attempting to attribute system behaviours to single components by disregarding that these components themselves are affected by the system as a whole.

6.3 Future work

While the tools developed in Chapter 3 have only been utilised in the analysis of two metabolic models, they have already demonstrated their utility for uncovering the mechanistic basis of the behaviour, control, and regulation of metabolic systems through model analysis. Thus, future application of these tools to other kinetic models of metabolic systems could similarly yield novel results. Specifically, we intend to utilise these tools in a comparative study of redoxin systems of different organisms. This will form part of a larger endeavour, which has already been initiated in a collaboration between our research group and that of Pillay (e.g. [193, 194, 232]), to elucidate the functional organisation of redoxin networks from a systems perspective, with a possible end goal of identifying drug targets within the redoxin network of *Mycobacterium tuberculosis*.

As discussed above, one of the strengths of implementing theoretical analysis frameworks as software tools is that it greatly enhances their potential for metabolic analysis through automation. One additional example of where automation could be implemented in our software is for identifying and comparing the backbone and multiplier components of control patterns. In Chapter 5 these patterns were identified using an *ad hoc*, partially-automated approach; i.e., while much of the identification and comparison process was carried out using software functions, these functions were developed specifically for the investigated system

and required some modification depending on the particular situation. Additionally, backbone patterns first had to be identified by hand before multiplier patterns could automatically be determined. In the future, this rather laborious task can be fully automated and included in the software as a standard (and generalised) functionality. These functions would have to identify backbone patterns based on the rules defined in Section 2.4.2, and subsequently determine the multiplier pattern from the expressions of the complete control patterns and the backbone patterns.

Finally, as many software developers will be able to attest, software is rarely ever truly finished. Thus, disregarding the implementation of additional features relating to methodological considerations, there will always be refinements and additions to be made in terms of the purely technical aspects of the software (not to mention the fixing of inevitable software "bugs"). One such refinement would be to remove the dependency of `PySCeSToolbox` on `Maxima` for equation simplification in `SymCa`, so that it would depend only on `SymPy` for symbolic computation. This would allow for easier distribution of the software, as it would require less additional software to be installed. At the present, however, this is not possible without a significant technical and time investment, as it would require substantial modifications to be made to `SymPy` itself (at least from the perspective of a non-expert in the field of computational algebra). Here we suspect that the central issue is the different factorisation algorithms implemented by each of these packages; `Maxima` uses Berlekamp's algorithm [233, 234], whereas `SymPy` uses an algorithm based on Kronecker's method [235, 236]. Currently, however, there are already efforts by the `SymPy` development team to implement a faster algorithm.

Another possible addition to `PySCeSToolbox` would be to extend its compatibility to other simulation packages. Currently, `PySCeSToolbox` depends significantly on the PySCeS API; however, in principle it could be modified to be compatible with any simulator with similar functionality as `PySCeS` that includes Python bindings or a Python API. This would mean that `PySCeSToolbox` would be able to rely on the simulation functionality of software such as the Systems Biology Workbench (`SBW`) [154], `Copasi` [237], `PyDStool` [238], or `ScrumPy` [150], as each of these software packages provide a Python API. There are many different possible approaches to implementing this functionality, however, it would most probably involve developing a common interface that would facilitate communication between `PySCeSToolbox` and the simulator. Such an interface could then be modified and extended to accommodate new simulators without the need to modify `PySCeSToolbox` itself. While implementing this functionality could potentially be technically challenging, it is an attractive prospect as it would enable users of other compatible simulation packages to simply "plug" `PySCeSToolbox` into their current workflow.

6.4 Conclusion

In this dissertation we have developed a software package `PySCeSToolbox` that implements theoretical analysis frameworks as three software tools. By using these tools to investigate previously published metabolic models, we could uncover previously undiscovered properties relating to their behaviour, control, and regulation. The greatest advantage of these tools was their ability to complement each other in model analysis, thus allowing us to arrive at a mechanistic explanation of a metabolic system. Ultimately, the work presented here affirms the utility of a marriage between different tools and approaches for understanding metabolic systems. As the field of systems biology advances and continues to grow, and the variety of tools at our disposal increases, it is important to utilise these tools to their full potential.

Appendices

Appendix A

Example Control Matrix Equation Formulation

Here follows an example of the construction of a control matrix equation for the pathway shown in Fig. 2.1 using the same procedure as previously demonstrated by Hofmeyr [57].

First we will repeat the 5×6-dimensional $\mathbf{N}$ originally shown in Equation 2.1, however, here it is augmented with an identity matrix where each column represents a time derivative.

	R1	R2	R3	R4	R5	R6	$\dot{s}_1$	$\dot{s}_2$	$\dot{s}_4$	$\dot{s}_5$	$\dot{s}_6$
S_1	1	−1	0	0	0	0	1	0	0	0	0
S_2	0	1	−1	−1	0	0	0	1	0	0	0
S_4	0	0	0	1	−1	0	0	0	1	0	0
S_5	0	0	0	0	1	−1	0	0	0	1	0
S_6	0	0	0	0	−1	1	0	0	0	0	1

(A.1)

Gaussian elimination is applied to the matrix above, transforming it to row echelon form and allowing us to determine the rank of the matrix:

	R1	R2	R3	R4	R5	R6	$\dot{s}_1$	$\dot{s}_2$	$\dot{s}_4$	$\dot{s}_5$	$\dot{s}_6$
S_1	1	−1	0	0	0	0	1	0	0	0	0
S_2	0	1	−1	−1	0	0	0	1	0	0	0
S_4	0	0	0	1	−1	0	0	0	1	0	0
S_5	0	0	0	0	1	−1	0	0	0	1	0
S_6	0	0	0	0	0	0	0	0	0	1	1

(A.2)

From the matrix above we can determine that $r = 2$. Therefore there are 4 independent fluxes and 2 independent metabolites. There is a single conservation relationship $\dot{s}_5 + \dot{s}_6 = T$, where T is the conserved sum.

As the columns of R3 and R6 in Equation A.1 do not have pivots, we choose J_3 and J_6 as the independent fluxes. The matrix $\mathbf{K}$ is constructed from the relationship $\mathbf{J} = \mathbf{K}\mathbf{J_i}$ as previously demonstrated in Equation 2.4:

$$\begin{bmatrix} J_3 \\ J_6 \\ J_1 \\ J_2 \\ J_4 \\ J_5 \end{bmatrix} = \begin{bmatrix} J_3 \\ J_6 \\ J_3+J_6 \\ J_3+J_6 \\ J_6 \\ J_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} J_3 + \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} J_6 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} J_3 \\ J_6 \end{bmatrix} \tag{A.3}$$

Now $\mathcal{K}$ is produced from the relationship $\mathcal{K} = (\mathbf{D^J})^{-1}\mathbf{K}\mathbf{D^{J_i}}$:

$$\begin{bmatrix} \frac{1}{J_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{J_6} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{J_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{J_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{J_4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{J_5} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} J_3 & 0 \\ 0 & J_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{J_3}{J_1} & \frac{J_6}{J_1} \\ \frac{J_3}{J_2} & \frac{J_6}{J_2} \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{A.4}$$

Note that the $(5,2)^{\text{th}}$ and $(6,2)^{\text{th}}$ entries of $\mathcal{K}$ are both 1 due to the flux relationship $J_6 = J_5 = J_4$.

Next, we construct $\mathbf{L}$. Here S_5 is chosen as the independent metabolite, therefore S_6 is the dependent metabolite. Thus, $\mathbf{L}$ is produced from the relationship $\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \mathbf{L}\frac{\mathrm{d}\mathbf{s_i}}{\mathrm{d}t}$:

$$\begin{bmatrix} \dot{s}_1 \\ \dot{s}_2 \\ \dot{s}_4 \\ \dot{s}_5 \\ \dot{s}_6 \end{bmatrix} = \begin{bmatrix} \dot{s}_1 \\ \dot{s}_2 \\ \dot{s}_4 \\ \dot{s}_5 \\ -\dot{s}_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \dot{s}_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \dot{s}_2 + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \dot{s}_4 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \dot{s}_5 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \dot{s}_1 \\ \dot{s}_2 \\ \dot{s}_4 \\ \dot{s}_5 \end{bmatrix} \tag{A.5}$$

$\mathcal{L}$ derives from the relationship $\mathcal{L} = (\mathbf{D^s})^{-1}\mathbf{L}\mathbf{D^{s_i}}$:

$$\begin{bmatrix} \frac{1}{s_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{s_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{s_4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{s_5} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{s_6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_4 & 0 \\ 0 & 0 & 0 & s_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{s_5}{s_6} \end{bmatrix} \tag{A.6}$$

The first step in constructing the matrix $\mathbf{E}$ is to produce the matrix $-\boldsymbol{\varepsilon}_{\mathbf{s}}\boldsymbol{\mathcal{L}}$:

$$
\begin{aligned}
-\begin{bmatrix}
0 & \varepsilon_{s_2}^{\nu_3} & 0 & 0 & 0 \\
0 & 0 & 0 & \varepsilon_{s_5}^{\nu_6} & \varepsilon_{s_6}^{\nu_6} \\
\varepsilon_{s_1}^{\nu_1} & 0 & 0 & 0 & 0 \\
\varepsilon_{s_1}^{\nu_2} & \varepsilon_{s_2}^{\nu_2} & 0 & 0 & 0 \\
0 & \varepsilon_{s_2}^{\nu_4} & \varepsilon_{s_4}^{\nu_4} & 0 & 0 \\
0 & 0 & \varepsilon_{s_4}^{\nu_5} & \varepsilon_{s_5}^{\nu_5} & \varepsilon_{s_6}^{\nu_5}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & -\frac{s_5}{s_6}
\end{bmatrix} \\
= \begin{bmatrix}
0 & -\varepsilon_{s_2}^{\nu_3} & 0 & 0 \\
0 & 0 & 0 & \left(\varepsilon_{s_6}^{\nu_6}\frac{s_5}{s_6} - \varepsilon_{s_5}^{\nu_6}\right) \\
-\varepsilon_{s_1}^{\nu_1} & 0 & 0 & 0 \\
-\varepsilon_{s_1}^{\nu_2} & -\varepsilon_{s_2}^{\nu_2} & 0 & 0 \\
0 & -\varepsilon_{s_2}^{\nu_4} & -\varepsilon_{s_4}^{\nu_4} & 0 \\
0 & 0 & -\varepsilon_{s_4}^{\nu_5} & \left(\varepsilon_{s_6}^{\nu_5}\frac{s_5}{s_6} - \varepsilon_{s_5}^{\nu_5}\right)
\end{bmatrix}
\end{aligned}
\tag{A.7}
$$

Finally, we can construct the control matrix equation for the independent variables $\mathbf{C}^{\mathbf{i}}\mathbf{E} = \mathbf{I}_n$ by augmenting $\boldsymbol{\mathcal{K}}$ with $-\boldsymbol{\varepsilon}_{\mathbf{s}}\boldsymbol{\mathcal{L}}$

$$
\begin{aligned}
&\begin{bmatrix}
C_3^{J_3} & C_6^{J_3} & C_1^{J_3} & C_2^{J_3} & C_4^{J_3} & C_5^{J_3} \\
C_3^{J_6} & C_6^{J_6} & C_1^{J_6} & C_2^{J_6} & C_4^{J_6} & C_5^{J_6} \\
C_3^{s_1} & C_6^{s_1} & C_1^{s_1} & C_2^{s_1} & C_4^{s_1} & C_5^{s_1} \\
C_3^{s_2} & C_6^{s_2} & C_1^{s_2} & C_2^{s_2} & C_4^{s_2} & C_5^{s_2} \\
C_3^{s_4} & C_6^{s_4} & C_1^{s_4} & C_2^{s_4} & C_4^{s_4} & C_5^{s_4} \\
C_3^{s_5} & C_6^{s_5} & C_1^{s_5} & C_2^{s_5} & C_4^{s_5} & C_5^{s_5}
\end{bmatrix}
\times
\begin{bmatrix}
1 & 0 & 0 & -\varepsilon_{s_2}^{\nu_3} & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & \left(\varepsilon_{s_6}^{\nu_6}\frac{s_5}{s_6} - \varepsilon_{s_5}^{\nu_6}\right) \\
\frac{J_3}{J_1} & \frac{J_6}{J_1} & -\varepsilon_{s_1}^{\nu_1} & 0 & 0 & 0 \\
\frac{J_3}{J_2} & \frac{J_6}{J_2} & -\varepsilon_{s_1}^{\nu_2} & -\varepsilon_{s_2}^{\nu_2} & 0 & 0 \\
0 & 1 & 0 & -\varepsilon_{s_2}^{\nu_4} & -\varepsilon_{s_4}^{\nu_4} & 0 \\
0 & 1 & 0 & 0 & -\varepsilon_{s_4}^{\nu_5} & \left(\varepsilon_{s_6}^{\nu_5}\frac{s_5}{s_6} - \varepsilon_{s_5}^{\nu_5}\right)
\end{bmatrix} \\
&= \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned}
\tag{A.8}
$$

Appendix B

Additional Tables

This appendix contains additional tables to Chapter 4 that were originally included as "Additional file 6". Table B.1 is equivalent to Table S1 contained within that file, and Table B.2 is equivalent to Table S2.

Table B.1: Metabolic control analysis of J_5, J_9, J_{14} and J_{15} for the aspartate metabolism model. Partial response coefficients and their associated elasticity and control coefficients are shown. Control analysis was performed with the concentration of linking metabolite between supply and demand fixed at the reference steady-state value in each case (see Table 4.2 in main text).

Elasticity coefficient		Control coefficient		Partial response coefficient	
ε^5_{ASA}	-2.517	C^5_5	0.029	${}^5R^5_{ASA}$	-0.072
ε^6_{ASA}	1.004	C^5_6	-0.275	${}^6R^5_{ASA}$	-0.276
ε^7_{ASA}	1.001	C^5_7	-0.320	${}^7R^5_{ASA}$	-0.320
ε^{15}_{ASA}	1.002	C^5_{15}	-0.337	${}^{15}R^5_{ASA}$	-0.338
ε^{14}_{ASA}	1.000	C^5_{14}	-0.037	${}^{14}R^5_{ASA}$	-0.037
ε^1_{Thr}	-2.357	C^9_1	0.015	${}^1R^9_{Thr}$	-0.035
ε^2_{Thr}	-1.763	C^9_2	0.078	${}^2R^9_{Thr}$	-0.137
ε^6_{Thr}	-0.327	C^9_6	0.193	${}^6R^9_{Thr}$	-0.063
ε^7_{Thr}	-0.025	C^9_7	0.224	${}^7R^9_{Thr}$	-0.006
ε^3_{Lys}	-0.857	C^{14}_3	0.113	${}^3R^{14}_{Lys}$	-0.097
ε^4_{Lys}	-0.833	C^{14}_4	0.534	${}^4R^{14}_{Lys}$	-0.445
ε^{15}_{Lys}	-1.630	C^{14}_{15}	-0.210	${}^{15}R^{14}_{Lys}$	0.342
ε^{14}_{Lys}	-1.956	C^{14}_{14}	0.977	${}^{14}R^{14}_{Lys}$	-1.911
ε^3_{Lys}	-0.857	C^{15}_3	0.113	${}^3R^{15}_{Lys}$	-0.097
ε^4_{Lys}	-0.833	C^{15}_4	0.534	${}^4R^{15}_{Lys}$	-0.445
ε^{15}_{Lys}	-1.630	C^{15}_{15}	0.790	${}^{15}R^{15}_{Lys}$	-1.288
ε^{14}_{Lys}	-1.956	C^{15}_{14}	-0.023	${}^{14}R^{15}_{Lys}$	0.045

Table B.2: Steady-state concentrations and fluxes for aspartate metabolism in the reference model and for knockouts. Rates are in $\mu M\ s^{-1}$ and concentrations are in μM).

	Reference	AKI knockout	AKII knockout	AKI,AKII knockout
ASA	0.96	0.95	0.88	0.84
Lys	69.16	68.69	65.99	64.74
Thr	296.93	285.74	228.99	206.81
J_1	0.018	-	0.033	-
J_2	0.095	0.102	-	-
J_3	0.158	0.159	0.165	0.169
J_4	0.745	0.750	0.780	0.794
J_5	1.016	1.011	0.978	0.962
J_6	0.324	0.323	0.320	0.317
J_7	0.377	0.372	0.346	0.334
J_{14}	0.031	0.031	0.031	0.031
J_{15}	0.284	0.284	0.280	0.279

Appendix C

Comparison Between the Fixed and Free Moiety-ratio Pyruvate Branch Models

The following figures serve to compare the results obtained for the fixed-NADH/NAD$^+$ model of Chapter 5 to those of the free-NADH/NAD$^+$ model. In Figs. C.1 and C.3, the value of V_{max}^{13} in the free-NADH/NAD$^+$ model was varied over a range to produce a range of ϕ_N values similar to that used to perform the analyses of Chapters 4 and 5. In Figs. C.2 and C.4, the results of Figs. C.1 and C.3 were scaled to this ϕ_N range. Note that while this represents a smaller range of ϕ_N values than used in Chapters 4 and 5, the general shapes of both the flux curve (Fig. C.2), and the control pattern contribution curve (Fig. C.4) bear an uncanny similarity to those of their fixed-NADH/NAD$^+$counterparts (Figs. 4.1A and 5.3A). Table C.1 gives the full expressions of the control pattern numerators of $C_{v_3}^{J_6}$ in the free-NADH/NAD$^+$ model.

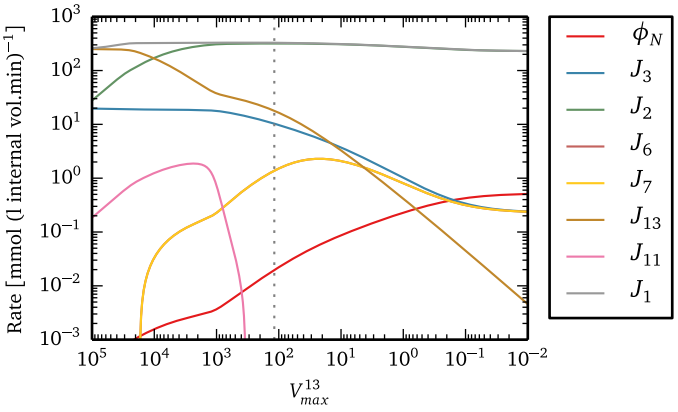


Figure C.1: The fluxes of the free-NADH/NAD$^+$ model, together with ϕ_N, as functions of V_{max}^{13}.

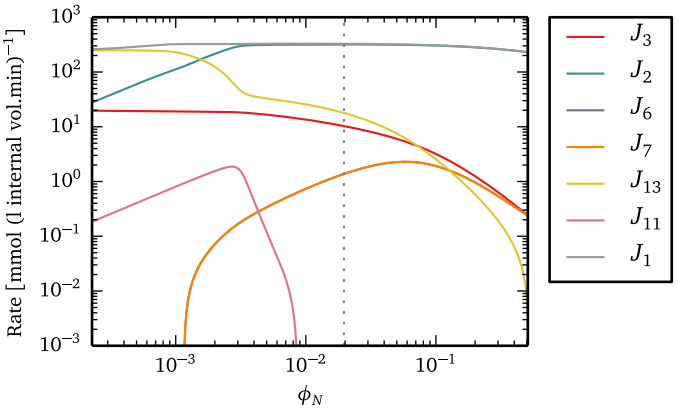


Figure C.2: The fluxes of the free-NADH/NAD$^+$ model as functions of ϕ_N.

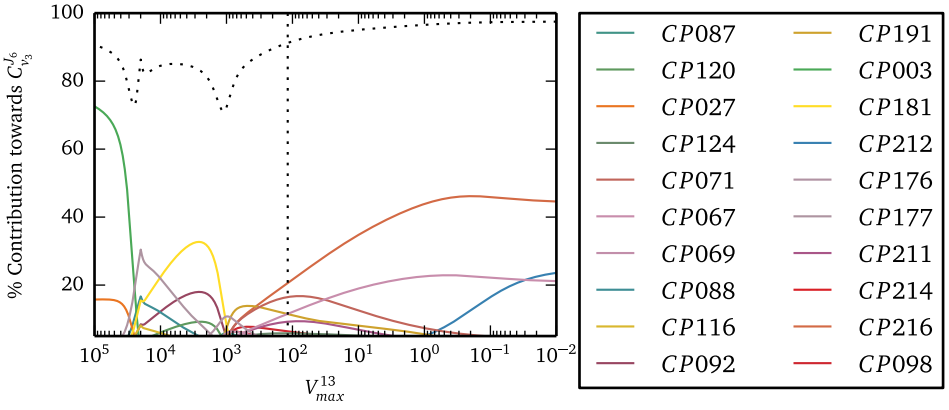


Figure C.3: The most important control patterns of $C^{J_6}_{v_3}$ as functions of V^{13}_{max} in the free-NADH/NAD$^+$ model

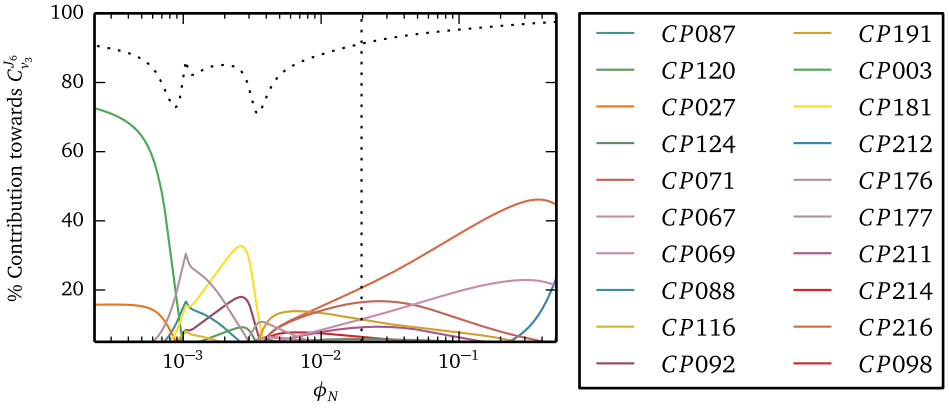


Figure C.4: The most important control patterns of $C^{J_6}_{v_3}$ as functions of ϕ_N in the free-NADH/NAD$^+$ model

Table C.1: Full numerator expressions of the dominant control patterns of $C_{\nu_3}^{J_6}$ in the unfixed model. Control patterns are arranged into groups and ordered as described in Table 5.4.

Group	Control Pattern	Expression
1	CP003	$J_1J_{10}J_{12}J_{13}J_3J_8\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Pyr}^{\nu_1}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_8}$
	CP027	$-J_1J_{10}J_{12}J_{13}J_3J_8\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Pyr}^{\nu_1}\varepsilon_{Acp}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_8}$
2A	CP177	$2J_{10}J_{12}J_{13}J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
	CP088	$-2J_1J_{10}J_{13}J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
	CP116	$-2J_{10}J_{12}J_{13}J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Acp}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
2B	CP181	$2J_{10}J_{12}J_2J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
	CP092	$-2J_1J_{10}J_2J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{\phi_N}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
	CP120	$-2J_{10}J_{12}J_2J_3J_8J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_2}\varepsilon_{Acp}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Pyr}^{\nu_8}\varepsilon_{Aclac}^{\nu_9}$
3	CP176	$2J_{10}J_{12}J_{13}J_2J_3J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP214	$4J_{10}J_{12}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP069	$-4J_{10}J_{12}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP087	$-2J_1J_{10}J_{13}J_2J_3J_9\varepsilon_{Acet}^{\nu_{10}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
4	CP216	$4J_{11}J_{12}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP067	$-4J_1J_{11}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP071	$-4J_{11}J_{12}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP212	$4J_1J_{11}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_N}^{\nu_1}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_4}\varepsilon_{\phi_A}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP191	$2J_{11}J_{12}J_{13}J_2J_3J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP211	$4J_1J_{11}J_2J_3J_4J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{\phi_N}^{\nu_4}\varepsilon_{Acp}^{\nu_5}\varepsilon_{Acal}^{\nu_6}\varepsilon_{\phi_N}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP098	$-2J_1J_{11}J_{13}J_2J_3J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{\phi_A}^{\nu_1}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_5}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$
	CP124	$-2J_{11}J_{12}J_{13}J_2J_3J_9\varepsilon_{Acet}^{\nu_{11}}\varepsilon_{\phi_A}^{\nu_{12}}\varepsilon_{\phi_N}^{\nu_{13}}\varepsilon_{Pyr}^{\nu_2}\varepsilon_{Acp}^{\nu_4}\varepsilon_{\phi_N}^{\nu_6}\varepsilon_{Acal}^{\nu_7}\varepsilon_{Aclac}^{\nu_9}$

Bibliography

[1] Hofmeyr, J.-H.S. (2007) The biochemical factory that autonomously fabricates itself: A systems biological view of the living cell, in *Systems Biology: Philosophical Foundations* (Boogerd, F., Bruggeman, F.J., Hofmeyr, J.-H.S. & Westerhoff, H.V., eds.), pp. 217–242, Elsevier, Amsterdam.

[2] Wouters, A.G. (2003) Four notions of biological function, *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 34*, 633–668.

[3] Fell, D. (2007) How can we understand metabolism?, in *Systems Biology: Philosophical Foundations* (Boogerd, F., Bruggeman, F.J., Hofmeyr, J.-H.S. & Westerhoff, H.V., eds.), 1 edn., pp. 87–101, Elsevier, Amsterdam.

[4] Mahner, M. & Bunge, M. (1997) *Foundations of Biophilosophy*, Springer, Berlin.

[5] Astbury, W.T. (1961) Molecular Biology or Ultrastructural Biology ?, *Nature 190*, 1124–1124.

[6] Alberghina, L. & Westerhoff, H. (2005) Systems biology: Did we know it all along?, in *Systems Biology: Definitions and Perspectives*, vol. 13 of *Topics in Current Genetics* (Alberghina, L. & Westerhoff, H., eds.), pp. 3–12, Springer-Verlag, Berlin/Heidelberg.

[7] Westerhoff, H.V. & Palsson, B.Ø. (2004) The evolution of molecular biology into systems biology, *Nat. Biotechnol. 22*, 1249–1252.

[8] Kacser, H. & Burns, J.A. (1973) The control of flux, *Symp. Soc. Exp. Biol. 27*, 65–104.

[9] Heinrich, R. & Rapoport, T.A. (1974) A linear steady-state treatment of enzymatic chains: General properties, control and effector strength, *Eur. J. Biochem. 42*, 89–95.

[10] Higgins, J. (1963) Analysis of Sequential Reactions, *Ann. N.Y. Acad. Sci. 108*, 305–321.

[11] Fell, D.A. (1992) Metabolic control analysis: a survey of its theoretical and experimental development., *Biochem. J. 286*, 313.

[12] Fell, D. (1996) *Understanding the control of metabolism*, no. 2 in Frontiers in metabolism, repr. edn., Portland Press, London.

[13] Heinrich, R. & Schuster, S. (1996) *The Regulation of Cellular Systems*, Chapman and Hall, New York.

[14] Hofmeyr, J.-H.S. & Cornish-Bowden, A. (2000) Regulating the cellular economy of supply and demand, *FEBS Lett. 476*, 47–51.

[15] Rohwer, J.M. & Hofmeyr, J.-H.S. (2008) Identifying and characterising regulatory metabolites with generalised supply-demand analysis, *J. Theor. Biol. 252*, 546–554.

[16] Rohwer, J.M. & Hofmeyr, J.-H.S. (2010) Kinetic and thermodynamic aspects of enzyme control and regulation, *J. Phys. Chem. B 114*, 16280–16289.

[17] Hofmeyr, J.-H.S. (1989) Control-pattern analysis of metabolic pathways, *Eur. J. Biochem. 186*, 343–354.

[18] Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival Jr., B., Assad-Garcia, N., Glass, J.I. & Covert, M.W. (2012) A whole-cell computational model predicts phenotype from genotype, *Cell 150*, 389–401.

[19] Fell, D.A. & Sauro, H.M. (1985) Metabolic control and its analysis. Additional relationships between elasticities and control coefficients, *Eur. J. Biochem. 148*, 555–561.

[20] Sauro, H.M., Small, J.R. & Fell, D.A. (1987) Metabolic control and its analysis. Extensions to the theory and matrix method, *Eur. J. Biochem. 165*, 215–221.

[21] Small, J.R. & Fell, D.A. (1989) The matrix method of metabolic control analysis: its validity for complex pathway structures, *J. Theor. Biol. 136*, 181–197.

[22] Thomas, S. & Fell, D.A. (1993) A computer program for the algebraic determination of control coefficients in Metabolic Control Analysis, *Biochem. J. 292*, 351–360.

[23] Rohwer, J.M., Akhurst, T.J. & Hofmeyr, J.-H.S. (2008) Symbolic control analysis of cellular systems, in *Experimental Standard Conditions of Enzyme Characterizations. Proceedings of the 3rd International Beilstein Workshop* (Hicks, M.G. & Kettner, C., eds.), pp. 137–148, Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt.

[24] Akhurst, T.J. (2011) *Symbolic control analysis of cellular systems*, Ph.D. thesis, Stellenbosch University.

[25] Hofmeyr, J.-H.S. (1995) Metabolic regulation: A control analytic perspective, *J. Bioenerg. Biomembr. 27*, 479–490.

[26] McCormick, A.J., Watt, D.A. & Cramer, M.D. (2009) Supply and demand: sink regulation of sugar accumulation in sugarcane, *J. Exp. Bot. 60*, 357–364.

[27] Thomas, S., Mooney, P.J.F., Burrell, M.M. & Fell, D.A. (1997) Metabolic Control Analysis of glycolysis in tuber tissue of potato (*Solanum tuberosum*): explanation for the low control coefficient of phosphofructokinase over respiratory flux, *Biochem. J. 322*, 119–127.

[28] Kroukamp, O., Rohwer, J., Hofmeyr, J.-H. & Snoep, J. (2002) Experimental supply-demand analysis of anaerobic yeast energy metabolism, *Mol. Biol. Rep. 29*, 203–209.

[29] Koebmann, B.J., Westerhoff, H.V., Snoep, J.L., Solem, C., Pedersen, M.B., Nilsson, D., Michelsen, O. & Jensen, P.R. (2002) The extent to which ATP demand controls the glycolytic flux depends strongly on the organism and conditions for growth, *Mol. Biol. Rep. 29*, 41–45.

[30] Yuan, J., Doucette, C.D., Fowler, W.U., Feng, X.-J., Piazza, M., Rabitz, H.A., Wingreen, N.S. & Rabinowitz, J.D. (2009) Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*, *Mol. Syst. Biol. 5*, 302.

[31] Christensen, C.D., Hofmeyr, J.-H.S. & Rohwer, J.M. (2015) Tracing regulatory routes in metabolism using generalised supply-demand analysis, *BMC Syst. Biol. 9*, 89.

[32] Michal, G. & Schomburg, D. (eds.) (2012) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, 2 edn., Wiley, Hoboken.

[33] Goldbeter, A. (1997) *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*, Cambridge University Press.

[34] Costa, R.S., Hartmann, A. & Vinga, S. (2016) Kinetic modeling of cell metabolism for microbial production, *J. Biotechnol. 219*, 126–141.

[35] Baghalian, K., Hajirezaei, M.-R. & Schreiber, F. (2014) Plant Metabolic Modeling: Achieving New Insight into Metabolism and Metabolic Engineering, *Plant Cell 26*, 3847–3866.

[36] Ramli, U.S., Tang, M., Quant, P.A., Guschina, I.A., Fawcett, T. & Harwood, J.L. (2014) Informed metabolic engineering of oil crops using control analysis, *Biocatalysis and Agricultural Biotechnology 3*, 49–52.

[37] Hofmeyr, J.-H.S. & Cornish-Bowden, A. (1991) Quantitative assessment of regulation in metabolic systems, *Eur. J. Biochem. 200*, 223–236.

[38] Reich, J. & Sel'kov, E. (1981) *Energy Metabolism Of The Cell: A Theoretical Treatise*, Academic Press, London.

[39] Meléndez-Hevia, E. & de Paz-Lugo, P. (2008) Branch-point stoichiometry can generate weak links in metabolism: the case of glycine biosynthesis, *J. Biosci. 33*, 771–780.

[40] Atkinson, D.E. (1977) *Cellular energy metabolism and its regulation*, Academic Press, New York.

[41] Meléndez-Hevia, E., Waddell, T.G. & Montero, F. (1994) Optimization of Metabolism: The Evolution of Metabolic Pathways Toward Simplicity Through the Game of the Pentose Phosphate Cycle, *J. Theor. Biol. 166*, 201–220.

[42] Monod, J., Changeux, J.-P. & Jacob, F. (1963) Allosteric proteins and cellular control systems, *J. Mol. Biol. 6*, 306–329.

[43] Monod, J., Wyman, J. & Changeux, J.P. (1965) On the nature of allosteric transitions: A plausible model, *J. Mol. Biol. 12*, 88–118.

[44] Cornish-Bowden, A. (2013) *Fundamentals of Enzyme Kinetics*, Wiley-Blackwell, Weinheim.

[45] Ma, H.-W. & Zeng, A.-P. (2003) The connectivity structure, giant strong component and centrality of metabolic networks, *Bioinformatics 19*, 1423–1430.

[46] Zhao, J., Tao, L., Yu, H., Luo, J., Cao, Z. & Li, Y. (2008) Bow-tie topological features of metabolic networks and the functional significance, *Chin. Sci. Bull. 52*, 1036–1045.

[47] Csete, M. & Doyle, J. (2004) Bow ties, metabolism and disease, *Trends Biotechnol. 22*, 446–450.

[48] Braess, P.-D.D.D. (1968) Über ein Paradoxon aus der Verkehrsplanung, *Unternehmensforschung 12*, 258–268.

[49] Schuster, S. & Fell, D. (2007) Modeling and Simulating Metabolic Networks, in *Bioinformatics-From Genomes to Therapies* (Lengauer, T., ed.), pp. 755–805, Wiley-Blackwell, Weinheim.

[50] Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S. & Palsson, B.Ø. (2002) Genome-Scale Metabolic Model of *Helicobacter pylori* 26695, *J. Bacteriol. 184*, 4582–4593.

[51] Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R. & Palsson, B.Ø. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data, *Proc. Natl. Acad. Sci. U.S.A. 104*, 1777–1782.

[52] Flahaut, N.A.L., Wiersma, A., Bunt, B.v.d., Martens, D.E., Schaap, P.J., Sijtsma, L., Santos, V.A.M.d. & Vos, W.M.d. (2013) Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation, *Appl. Microbiol. Biotechnol. 97*, 8729–8739.

[53] Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S. & Palsson, B.Ø. (2004) Comparison of network-based pathway analysis methods, *Trends Biotechnol. 22*, 400–405.

[54] Schallau, K. & Junker, B.H. (2010) Simulating plant metabolic pathways with enzyme-kinetic models, *Plant Physiol. 152*, 1763–1771.

[55] Rohwer, J.M. (2012) Kinetic modelling of plant metabolic pathways, *J. Exp. Bot. 63*, 2275–2292.

[56] Reder, C. (1988) Metabolic control theory: A structural approach, *J. Theor. Biol. 135*, 175–201.

[57] Hofmeyr, J.-H.S. (2001) Metabolic control analysis in a nutshell, in *Proceedings of the 2nd International Conference on Systems Biology* (Yi, T., Hucka, M., Morohashi, M. & Kitano, H., eds.), pp. 291–300, Omnipress, Madison.

[58] Schuster, S., Dandekar, T. & Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering, *Trends Biotechnol. 17*, 53–60.

[59] Schuster, S. & Hilgetag, C. (1994) On elementary flux modes in biochemical reaction systems at steady state, *J. Biol. Syst. 02*, 165–182.

[60] Schuster, S., Fell, D.A. & Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nat. Biotechnol. 18*, 326–332.

[61] Schilling, C.H., Letscher, D. & Palsson, B.Ø. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective, *J. Theor. Biol. 203*, 229–248.

[62] Klamt, S. & Stelling, J. (2003) Two approaches for metabolic pathway analysis?, *Trends Biotechnol. 21*, 64–69.

[63] Schilling, C.H., Edwards, J.S. & Palsson, B.Ø. (1999) Toward Metabolic Phenomics: Analysis of Genomic Data Using Flux Balances, *Biotechnol. Prog. 15*, 288–295.

[64] Schilling, C.H., Edwards, J.S., Letscher, D. & Palsson, B.Ø. (2000) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems, *Biotechnol. Bioeng. 71*, 286–306.

[65] Orth, J.D., Thiele, I. & Palsson, B.Ø. (2010) What is flux balance analysis?, *Nat. Biotechnol. 28*, 245–248.

[66] Pfau, T., Christian, N. & Ebenhöh, O. (2011) Systems approaches to modelling pathways and networks, *Brief. Funct. Genomics 10*, 266–279.

[67] Kelk, S.M., Olivier, B.G., Stougie, L. & Bruggeman, F.J. (2012) Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks, *Sci. Rep. 2*, 580.

[68] Edwards, J.S. & Palsson, B.Ø. (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities, *Proc. Natl. Acad. Sci. U.S.A. 97*, 5528–5533.

[69] Schuetz, R., Kuepfer, L. & Sauer, U. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*, *Mol. Syst. Biol. 3*, 119.

[70] Varma, A. & Palsson, B.Ø. (1993) Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors, *J. Theor. Biol. 165*, 477–502.

[71] Oberhardt, M.A., Palsson, B.Ø. & Papin, J.A. (2009) Applications of genome-scale metabolic reconstructions, *Mol. Syst. Biol. 5*, 320.

[72] Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A. & Palsson, B.Ø. (2003) Metabolic pathways in the post-genome era, *Trends Biochem. Sci. 28*, 250–258.

[73] Bruggeman, F.J., Bakker, B.M., Hornberg, J.J. & Westerhoff, H.V. (2006) Introduction to Computational Models of Biochemical Reaction Networks, in *Computational Systems Biology* (Kriete, A. & Eils, R., eds.), pp. 127–148, Academic Press, Burlington.

[74] Gianchandani, E.P., Oberhardt, M.A., Burgard, A.P., Maranas, C.D. & Papin, J.A. (2008) Predicting biological system objectives *de novo* from internal state measurements, *BMC Bioinformatics 9*, 43.

[75] Knorr, A.L., Jain, R. & Srivastava, R. (2007) Bayesian-based selection of metabolic objective functions, *Bioinformatics 23*, 351–357.

[76] Burgard, A.P. & Maranas, C.D. (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions, *Biotechnol. Bioeng. 82*, 670–677.

[77] Michaelis, L. & Menten, M.L. (1913) Die Kinetik der Invertinwirkung, *Biochem. Z. 49*, 352.

[78] Adair, G.S., Bock, A.V. & Field, J.H. (1925) The hemoglobin system VI. The oxygen dissociation curve of hemoglobin, *J. Biol. Chem. 63*, 529–545.

[79] Hill, A.V. (1910) The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves, *The Journal of Physiology 40*, i–vii.

[80] Hofmeyr, J.-H.S. & Cornish-Bowden, H. (1997) The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models, *Computer applications in the biosciences : CABIOS 13*, 377–385.

[81] Popova, S. & Sel'kov, E. (1975) Generalization of the model by Monod, Wyman and Changeux for the case of a reversible monosubstrate reaction, *FEBS Lett. 53*, 269–273.

[82] Rohwer, J., Hanekom, A., Crous, C., Snoep, J. & Hofmeyr, J.-H. (2006) Evaluation of a simplified generic bi-substrate rate equation for computational systems biology, *Systems Biology, IEE Proceedings 153*, 338–341.

[83] Rohwer, J., Hanekom, A. & Hofmeyr, J.-H. (2006) A Universal Rate Equation for Systems Biology, in *Experimental Standard Conditions of Enzyme Characterizations. Proceedings of the 2nd International Beilstein Workshop* (Hicks, M.G. & Kettner, C., eds.), pp. 175–187, Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt.

[84] Rohwer, J. & Botha, F. (2001) Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data., *Biochem. J. 358*, 437–445.

[85] Curien, G., Bastien, O., Robert-Genthon, M., Cornish-Bowden, A., Cárdenas, M.L. & Dumas, R. (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters, *Mol. Syst. Biol. 5*.

[86] Hoefnagel, M.H.N., Starrenburg, M.J.C., Martens, D.E., Hugenholtz, J., Kleerebezem, M., Swam, I.I.V., Bongers, R., Westerhoff, H.V. & Snoep, J.L. (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis, *Microbiology 148*, 1003–1013.

[87] Galazzo, J.L. & Bailey, J.E. (1990) Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*, *Enzyme Microb. Technol. 12*, 162–172.

[88] The Sage Developers (2016) *SageMath, the Sage Mathematics Software System (Version 7.0)*, http://www.sagemath.org.

[89] The MathWorks Inc. (2016) *MATLAB version 9.1 (2016b)*, https://www.mathworks.com/.

[90] Wolfram, S. (1991) *Mathematica: A System for Doing Mathematics by Computer (2Nd Ed.)*, Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

[91] Olivier, B.G., Rohwer, J.M. & Hofmeyr, J.H.S. (2004) Modelling cellular systems with PySCeS, *Bioinformatics 21*, 560–561.

[92] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. & Kummer, U. (2006) Copasi - a complex pathway simulator, *Bioinformatics 22*, 3067–3074.

[93] Hindmarsh, A.C. (1983) Odepack, a systematized collection of ode solvers, in *Scientific Computing*, vol. 1 of *IMACS Transactions on scientific computing* (Stepleman, R., ed.), pp. 55–64, Elsevier, Amsterdam.

[94] Petzold, L. (1983) Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations, *SIAM Journal on Scientific and Statistical Computing 4*, 136–148.

[95] Deuflhard, P. (2011) *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, no. 35 in Springer Series in Computational Mathematics, 1 edn., Springer, Berlin.

[96] Burns, J.A., Cornish-Bowden, A., Groen, A.K., Heinrich, R., Kacser, H., Porteous, J.W., Rapoport, S.M., Rapoport, T.A., Stucki, J.W., Tager, J.M., Wanders, R.J.A. & Westerhoff, H.V. (1985) Control analysis of metabolic systems, *Trends Biochem. Sci. 10*, 16.

[97] Westerhoff, H.V. & Chen, Y.-D. (1984) How do enzyme activities control metabolite concentrations?. An additional theorem in the theory of metabolic control, *Eur. J. Biochem. 142*, 425–430.

[98] Westerhoff, H.V. & Kell, D.B. (1987) Matrix method for determining steps most rate–limiting to metabolic fluxes in biotechnological processes, *Biotechnol. Bioeng. 30*, 101–107.

[99] Giersch, C. (1988) Control analysis of metabolic networks. 1. Homogeneous functions and the summation theorems for control coefficients, *Eur. J. Biochem. 174*, 509–513.

[100] Giersch, C. (1988) Control analysis of metabolic networks. 2. Total differentials and general formulation of the connectivity relations, *Eur. J. Biochem. 174*, 515–519.

[101] Giersch, C. (1988) Control analysis of biochemical pathways: A novel procedure for calculating control coefficients, and an additional theorem for branched pathways, *J. Theor. Biol. 134*, 451–462.

[102] Cascante, M., Franco, R. & Canela, E.I. (1989) Use of implicit methods from general sensitivity theory to develop a systematic approach to metabolic control. I. unbranched pathways, *Math. Biosci. 94*, 271–288.

[103] Cascante, M., Franco, R. & Canela, E.I. (1989) Use of implicit methods from general sensitivity theory to develop a systematic approach to metabolic control. II. complex systems, *Math. Biosci. 94*, 289–309.

[104] Fell, D.A., Sauro, H.M. & Small, J.R. (1990) Control Coefficients and the Matrix Method, in *Control of Metabolic Processes* (Cornish-Bowden, A. & Cárdenas, M.L., eds.), no. 190 in NATO ASI Series, pp. 139–148, Springer, New York.

[105] Kacser, H., Sauro, H.M. & Acerenza, L. (1990) Enzyme-enzyme interactions and control analysis, *Eur. J. Biochem. 187*, 481–491.

[106] Sauro, H.M. & Kacser, H. (1990) Enzyme-enzyme interactions and control analysis, *Eur. J. Biochem. 187*, 493–500.

[107] Hofmeyr, J.-H., Cornish-Bowden, A. & Rohwer, J.M. (1993) Taking enzyme kinetics out of control; putting control into regulation, *Eur. J. Biochem. 212*, 833–837.

[108] van der Gugten, A.A. & Westerhoff, H.V. (1993) Internal Regulation. The C·E = I = E·C Square-Matrix Method Illustrated for a Simple Case of a Complex Pathway, in *Modern Trends in Biothermokinetics* (Schuster, S., Rigoulet, M., Ouhabi, R. & Mazat, J.-P., eds.), pp. 253–262, Springer, Boston.

[109] Westerhoff, H.V., Hofmeyr, J.-H.S. & Kholodenko, B.N. (1994) Getting to the inside of cells using metabolic control analysis, *Biophys. Chem. 50*, 273–283.

[110] Hofmeyr, J.-H.S. & Cornish-Bowden, A. (1996) Co-response analysis: A new experimental strategy for metabolic control analysis, *J. Theor. Biol. 182*, 371–380.

[111] Savageau, M. (1987) Control of metabolism: where is the theory?, *Trends Biochem. Sci. 12*, 219–220.

[112] Savageau, M.A., Voit, E.O. & Irvine, D.H. (1987) Biochemical systems theory and metabolic control theory: 1. fundamental similarities and differences, *Math. Biosci. 86*, 127–145.

[113] Savageau, M.A., Voit, E.O. & Irvine, D.H. (1987) Biochemical systems theory and metabolic control theory: 2. the role of summation and connectivity relationships, *Math. Biosci. 86*, 147–169.

[114] Sauro, H.M. & Fell, D.A. (1991) SCAMP: A metabolic simulator and control analysis program, *Mathematical and Computer Modelling 15*, 15–28.

[115] Hofmeyr, J.H.S. & van der Merwe, K.J. (1986) METAMOD: software for steady-state modelling and control analysis of metabolic pathways on the BBC microcomputer, *Computer applications in the biosciences : CABIOS 2*, 243–249.

[116] Higham, N.J. (2002) *Accuracy and stability of numerical algorithms*, 2 edn., Society for Industrial and Applied Mathematics, Philadelphia.

[117] Mendes, P. (1993) GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems, *Computer applications in the biosciences : CABIOS 9*, 563–571.

[118] Hofmeyr, J.-H.S. (1986) *Studies in Steady-State Modelling And Control Analysis of Metabolic Systems*, Ph.D. thesis, Stellenbosch University.

[119] Hofmeyr, J.-H.S. (1990) Control-Pattern Analysis of Metabolic Systems, in *Control of Metabolic Processes* (Cornish-Bowden, A. & Cárdenas, M.L., eds.), no. 190 in NATO ASI Series, pp. 239–248, Springe, New York.

[120] Schulz, A.R. (1991) Algorithms for the derivation of Flux and Concentration Control Coefficients., *Biochem. J. 278*, 299–304.

[121] Hofmeyr, J.H.S. (1986) Steady-state modelling of metabolic pathways: A guide for the prospective simulator, *Computer applications in the biosciences : CABIOS 2*, 5–11.

[122] Thomas, S. & Fell, D.A. (1998) A control analysis exploration of the role of ATP utilisation in glycolytic-flux control and glycolytic-metabolite-concentration regulation, *Eur. J. Biochem. 258*, 956–967.

[123] Uys, L., Botha, F.C., Hofmeyr, J.-H.S. & Rohwer, J.M. (2007) Kinetic model of sucrose accumulation in maturing sugarcane culm tissue, *Phytochemistry 68*, 2375–2392.

[124] Kacser, H. & Burns, J. (1979) Molecular democracy: who shares the controls?, *Biochem. Soc. Trans. 7*, 1149–1160.

[125] Koebmann, B.J., Westerhoff, H.V., Snoep, J.L., Nilsson, D. & Jensen, P.R. (2002) The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP, *J. Bacteriol. 184*, 3909–3916.

[126] du Preez, F., Conradie, R., Penkler, G., Holm, K., van Dooren, F. & Snoep, J. (2008) A comparative analysis of kinetic models of erythrocyte glycolysis, *J. Theor. Biol. 252*, 488–496.

[127] Jørgensen, C.M., Hammer, K., Jensen, P.R. & Martinussen, J. (2004) Expression of the pyrG gene determines the pool sizes of CTP and dCTP in *Lactococcus lactis*, *Eur. J. Biochem. 271*, 2438–2445.

[128] Arsac, L.M., Nouette-Gaulain, K., Miraux, S., Deschodt-Arsac, V., Rossignol, R., Thiaudiere, E. & Diolez, P. (2012) Acute and chronic effects of bupivacaine on muscle energetics during contraction in vivo: a modular metabolic control analysis, *Biochem. J. 444*, 315–321.

[129] Morandini, P. (2013) Control limits for accumulation of plant metabolites: brute force is no substitute for understanding, *Plant Biotechnol. J. 11*, 253–267.

[130] Roesler, K., Shintani, D., Savage, L., Boddupalli, S. & Ohlrogge, J. (1997) Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds, *Plant Physiology 113*, 75–81.

[131] Madoka, Y., Tomizawa, K.-I., Mizoi, J., Nishida, I., Nagano, Y. & Sasaki, Y. (2002) Chloroplast transformation with modified accD operon increases acetyl-CoA carboxylase and causes extension of leaf longevity and increase in seed yield in tobacco, *Plant & Cell Physiology 43*, 1518–1525.

[132] Bouvier-Navé, P., Benveniste, P., Oelkers, P., Sturley, S.L. & Schaller, H. (2000) Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase, *European journal of biochemistry / FEBS 267*, 85–96.

[133] Oakes, J., Brackenridge, D., Colletti, R., Daley, M., Hawkins, D.J., Xiong, H., Mai, J., Screen, S.E., Val, D., Lardizabal, K., Gruys, K. & Deikman, J. (2011) Expression of fungal diacylglycerol acyltransferase2 genes to increase kernel oil in maize, *Plant Physiology 155*, 1146–1157.

[134] Schafer, J.R.A., Fell, D.A., Rothman, D. & Shulman, R.G. (2004) Protein phosphorylation can regulate metabolite concentrations rather than control flux: The example of glycogen synthase, *Proc. Natl. Acad. Sci. U.S.A. 101*, 1485–1490.

[135] Rolleston, F.S. (1972) A theoretical background to the use of measured concentrations of intermediates in study of the control of intermediary metabolism, in *Current Topics in Cellular Regulation*, vol. 5 (Horecker, B.L. & Stadtman, E.R., eds.), pp. 47–75, Academic Press, New York.

[136] Morandini, P. (2009) Rethinking metabolic control, *Plant Science 176*, 441–451.

[137] Chance, B. (1943) The Kinetics of the Enzyme-Substrate Compound of Peroxidase, *J. Biol. Chem. 151*, 553–577.

[138] Chance, B., Greenstein, D.S., Higgins, J. & Yang, C.C. (1952) The mechanism of catalase action. II. Electric analog computer studies, *Arch. Biochem. Biophys. 37*, 322–339.

[139] Chance, B. (1960) Analogue and digital representations of enzyme kinetics, *The Journal of Biological Chemistry 235*, 2440–2443.

[140] Chance, B., Garfinkel, D., Higgins, J. & Hess, B. (1960) Metabolic control mechanisms. 5. A solution for the equations representing interaction between glycolysis and respiration in ascites tumor cells, *The Journal of Biological Chemistry 235*, 2426–2439.

[141] Garfinkel, D., Garfinkel, L., Pring, M., Green, S.B. & Chance., B. (1970) Computer Applications to Biochemical Kinetics, *Annu. Rev. Biochem. 39*, 473–498.

[142] SBML Team (2016), The SBML Software Guide, http://sbml.org/SBML_Software_Guide.

[143] Cornish-Bowden, A. & Hofmeyr, J.H. (1991) MetaModel: a program for modelling and control analysis of metabolic pathways on the IBM PC and compatibles, *Computer applications in the biosciences: CABIOS 7*, 89–93.

[144] Letellier, T., Reder, C. & Mazat, J.-P. (1991) CONTROL: software for the analysis of the control of metabolic networks, *Computer applications in the biosciences : CABIOS 7*, 383–390.

[145] Sauro, H.M. (1993) SCAMP: a general-purpose simulator and metabolic control analysis program, *Computer applications in the biosciences : CABIOS 9*, 441–450.

[146] Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3, *Trends Biochem. Sci. 22*, 361–363.

[147] Ehlde, M. & Zacchi, G. (1995) MIST: a user-friendly metabolic simulator, *Computer applications in the biosciences : CABIOS 11*, 201–207.

[148] Goryanin, I., Hodgman, T.C. & Selkov, E. (1999) Mathematical simulation and analysis of cellular metabolism and regulation., *Bioinformatics 15*, 749–758.

[149] Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks, *BIOSILICO 1*, 159–162.

[150] Poolman, M. (2006) Scrumpy: Metabolic modelling with Python, *IEE Proceedings: Systems Biology 153*, 375–378.

[151] Moraru, I., Morgan, F., Li, Y., Loew, L., Schaff, J., Lakshminarayana, A., Slepchenko, B., Gao, F. & Blinov, M. (2008) Virtual Cell modelling and simulation software environment, *IET Syst. Biol. 2*, 352–362.

[152] Resasco, D.C., Gao, F., Morgan, F., Novak, I.L., Schaff, J.C. & Slepchenko, B.M. (2012) Virtual Cell: computational tools for modeling in cell biology, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine 4*, 129–140.

[153] Sauro, H.M. (2000) Jarnac: a system for interactive metabolic analysis, in *Animating the Cellular Map 9th International BioThermoKinetics Meeting* (Hofmeyr, J.-H.S., Rohwer, J.M. & Snoep, J.L., eds.), no. 33 in Animating the Cellular Map 9th International BioThermoKinetics Meeting, pp. 221–228, Stellenbosch University Press.

[154] Bergmann, F.T. & Sauro, H.M. (2006) Sbw - a modular framework for systems biology, in *Proceedings of the 38th Conference on Winter Simulation*, pp. 1637–1645, Winter Simulation Conference, Monterey.

[155] Olivier, B.G., Rohwer, J.M. & Hofmeyr, J.-H.S. (2002) Modelling cellular processes with Python and SciPy, *Mol. Biol. Rep. 29*, 249–254.

[156] Jones, E., Oliphant, T., Peterson, P. *et al.* (2001–), SciPy: Open source scientific tools for Python, http://www.scipy.org/.

[157] Oliphant, T.E. (2007) Python for scientific computing, *Computing in Science Engineering 9*, 10–20.

[158] Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Forum, a.t.r.o.t.S., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Novère, N.L., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J. & Wang, J. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics 19*, 524–531.

[159] Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L. & Hucka, M. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, *Nucleic Acids Res. 34*, D689–D691.

[160] Olivier, B.G. & Snoep, J.L. (2004) Web-based kinetic modelling using JWS online, *Bioinformatics 20*, 2143–2144.

[161] Molloy, J.C. (2011) The Open Knowledge Foundation: Open Data Means Better Science, *PLoS Biol. 9*, e1001195.

[162] Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I. & Müller,

W. (2012) SABIO-RK—database for biochemical reaction kinetics, *Nucleic Acids Res. 40*, D790–D796.

[163] Chang, A., Scheer, M., Grote, A., Schomburg, I. & Schomburg, D. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009, *Nucleic Acids Res. 37*, D588–D592.

[164] Kanehisa, M. & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res. 28*, 27–30.

[165] Somogyi, E.T., Bouteiller, J.-M., Glazier, J.A., König, M., Medley, J.K., Swat, M.H. & Sauro, H.M. (2015) libRoadRunner: a high performance SBML simulation and analysis library, *Bioinformatics 31*, 3315–3321.

[166] Cascante, M., Boros, L.G., Comin-Anduix, B., de Atauri, P., Centelles, J.J. & Lee, P.W.-N. (2002) Metabolic control analysis in drug discovery and disease, *Nat. Biotechnol. 20*, 243–249.

[167] Mardinoglu, A. & Nielsen, J. (2015) New paradigms for metabolic modeling of human cells, *Curr. Opin. Biotechnol. 34*, 91–97.

[168] Chakrabarti, A., Miskovic, L., Soh, K.C. & Hatzimanikatis, V. (2013) Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints, *Biotechnol. J. 8*, 1043–1057.

[169] Shi, H. & Schwender, J. (2016) Mathematical models of plant metabolism, *Curr. Opin. Biotechnol. 37*, 143–152.

[170] Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2014) *Julia: A Fresh Approach to Numerical Computing*, http://arxiv.org/abs/1411.1607.

[171] Glazebrook, K., Brinchmann, J., Cerney, J., DeForest, C., Hunt, D., Jenness, T., Luka, T., Schwebel, R. & Soeller, C. (1997) The perl data language, *The Perl Journal 5*.

[172] Jupyter Steering Council (2016), Project Jupyter, http://jupyter.org/.

[173] Pérez, F. & Granger, B. (2007) IPython: A System for Interactive Scientific Computing, *Computing in Science Engineering 9*, 21–29.

[174] Hunter, J.D. (2007) Matplotlib: A 2D graphics environment, *Computing In Science & Engineering 9*, 90–95.

[175] SymPy Development Team (2012) *SymPy: Python library for symbolic mathematics*, http://www.sympy.org.

[176] Maxima (2016), Maxima, a computer algebra system. version 5.38.1, http://maxima.sourceforge.net/.

[177] D3.js Developers (2016), D3.js - Data Driven Documents, https://d3js.org/.

[178] NetworkX Developers (2016), NetworkX, http://networkx.github.io/.

[179] van der Walt, S., Colbert, S. & Varoquaux, G. (2011) The NumPy Array: A Structure for Efficient Numerical Computation, *Computing in Science Engineering 13*, 22–30.

[180] Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A. & Milo, R. (2013) A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects, *FEBS Lett. 587*, 2772–2777.

[181] Millman, K.J. & Aivazis, M. (2011) Python for Scientists and Engineers, *Computing in Science Engineering 13*, 9–12.

[182] Perez, F., Granger, B.E. & Hunter, J.D. (2011) Python: An Ecosystem for Scientific Computing, *Computing in Science Engineering 13*, 13–21.

[183] R Core Team (2013) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

[184] John W. Eaton, David Bateman, S.H. & Wehbring, R. (2014) *GNU Octave version 3.8.1 manual: a high-level interactive language for numerical computations*, CreateSpace Independent Publishing Platform.

[185] Prechelt, L. (2000) An empirical comparison of seven programming languages, *Computer 33*, 23–29.

[186] Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D.S. & Smith, K. (2011) Cython: The Best of Both Worlds, *Computing in Science Engineering 13*, 31–39.

[187] Teusink, B. & Westerhoff, H.V. (2000) 'Slave' metabolites and enzymes *267*, 1889–1893.

[188] Sauro, H.M. (1994) Moiety-conserved cycles and metabolic control analysis: problems in sequestration and metabolic channelling, *Biosystems 33*, 55–67.

[189] Thomas, S. & Fell, D.A. (1996) Design of Metabolic Control for Large Flux Changes, *J. Theor. Biol. 182*, 285–298.

[190] Rohwer, J.M., Viljoen, C., Christensen, C.D., Mashamaite, L.N. & Pillay, C.S. (2016) Identifying the conditions necessary for the thioredoxin ultrasensitive response, *Perspectives in Science doi:10.1016/j.pisc.2016.05.011*, in press.

[191] Liebermeister, W., Uhlendorf, J. & Klipp, E. (2010) Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation, *Bioinformatics 26*, 1528–1534.

[192] Hoefnagel, M.H.N., Burgt, A.v.d., Martens, D.E., Hugenholtz, J. & Snoep, J.L. (2002) Time dependent responses of glycolytic intermediates in a detailed glycolytic model of *Lactococcus lactis* during glucose run-out experiments, *Mol. Biol. Rep. 29*, 157–161.

[193] Pillay, C.S., Hofmeyr, J.-H.S. & Rohwer, J.M. (2011) The logic of kinetic regulation in the thioredoxin system, *BMC Syst. Biol. 5*, 15.

[194] Pillay, C.S., Hofmeyr, J.-H., Mashamaite, L.N. & Rohwer, J.M. (2012) From Top-Down to Bottom-Up: Computational Modeling Approaches for Cellular Redoxin Networks, *Antioxidants & Redox Signaling 18*, 2075–2086.

[195] Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M. & Palsson, B.Ø. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism, *Mol. Syst. Biol. 7*, 535.

[196] Thiele, I., Jamshidi, N., Fleming, R.M.T. & Palsson, B.Ø. (2009) Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization, *PLoS Comput Biol 5*, e1000312.

[197] Lloyd, C.M., Lawson, J.R., Hunter, P.J. & Nielsen, P.F. (2008) The CellML model repository, *Bioinformatics 24*, 2122–2123.

[198] Snoep, J.L. & Westerhoff, H.V. (2005) From isolation to integration, a systems biology approach for building the silicon cell, in *Systems Biology* (Alberghina, L. & Westerhoff, H.V., eds.), no. 13 in Topics in Current Genetics, pp. 13–30, Springer Berlin Heidelberg.

[199] Kacser, H., Burns, J. & Fell, D. (1995) The control of flux: 21 years on, *Biochem. Soc. Trans. 23*, 341–366.

[200] Hofmeyr, J.-H.S. & Rohwer, J.M. (2011) Supply-demand analysis: A framework for exploring the regulatory design of metabolism, in *Methods in Enzymology*, vol. 500 (Jameson, D., Verma, M. & Westerhoff, H.V., eds.), pp. 533–554, Elsevier, Oxford.

[201] Olivier, B., Rohwer, J. & Hofmeyr, J. (2016) *PySCeS User Guide*, http://pysces.sourceforge.net/docs/index.html.

[202] Hucka, M., Bergmann, F.T., Keating, S.M. & Smith, L.P. (2011) A Profile of Today's SBML-Compatible Software, in *Seventh IEEE International Conference on e-Science Workshops*, pp. 143–150, Stockholm.

[203] Curien, G., Dumas, R., Cornish-Bowden, A. & Cardenas, M.L. (2010) Different contributions of the various isoenzymes to the flux in the aspartate-derived amino acid pathway in *Arabidopsis thaliana*, in *Proceedings of the 4th International Beilstein Symposium on Experimental Standard Conditions of Enzyme Characterizations* (Hicks, M.G. & Kettner, C., eds.), pp. 107–121, Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt.

[204] Kowalczyk, M. & Bardowski, J. (2007) Regulation of sugar catabolism in *Lactococcus lactis*, *Crit. Rev. Microbiol. 33*, 1–13.

[205] Levering, J., Musters, M.W.J.M., Bekker, M., Bellomo, D., Fiedler, T., de Vos, W.M., Hugenholtz, J., Kreikemeyer, B., Kummer, U. & Teusink, B. (2012) Role of phosphate in the central metabolism of two lactic acid bacteria – a comparative systems biology approach, *FEBS J. 279*, 1274–1290.

[206] Oh, E., Lu, M., Park, C., Park, C., Oh, H.B., Lee, S.Y. & Lee, J. (2011) Dynamic modeling of lactic acid fermentation metabolism with *Lactococcus lactis*, *J. Microbiol. Biotechnol. 21*, 162–169.

[207] Costa, R.S., Hartmann, A., Gaspar, P., Neves, A.R. & Vinga, S. (2014) An extended dynamic model of *Lactococcus lactis* metabolism for mannitol and 2,3-butanediol production, *Mol. BioSyst. 10*, 628–639.

[208] Palmfeldt, J., Paese, M., Hahn-Hägerdal, B. & Niel, E.W.J.v. (2004) The pool of ADP and ATP regulates anaerobic product formation in resting cells of *Lactococcus lactis*, *Appl. Environ. Microbiol. 70*, 5477–5484.

[209] Cao, R., Zeidan, A.A., Rådström, P. & van Niel, E.W.J. (2010) Inhibition kinetics of catabolic dehydrogenases by elevated moieties of ATP and ADP – implication for a new regulation mechanism in *Lactococcus lactis*, *FEBS Journal 277*, 1843–1852.

[210] Garrigues, C., Loubiere, P., Lindley, N.D. & Cocaign-Bousquet, M. (1997) Control of the shift from homolactic acid to mixed-acid fermentation in *Lactococcus lactis*: predominant role of the NADH/NAD^+ ratio., *J. Bacteriol. 179*, 5282–5287.

[211] van Hoek, M.J. & Merks, R.M. (2012) Redox balance is key to explaining full vs. partial switching to low-yield metabolism, *BMC Systems Biology 6*, 22.

[212] Lopez de Felipe, F., Starrenburg, M.J. & Hugenholtz, J. (1997) The role of NADH-oxidation in acetoin and diacetyl production from glucose in *Lactococcus lactis* subsp. *lactis* MG1363, *FEMS Microbiology Letters 156*, 15–19.

[213] Hols, P., Ramos, A., Hugenholtz, J., Delcour, J., Vos, W.M.d., Santos, H. & Kleerebezem, M. (1999) Acetate utilization in *Lactococcus lactis* deficient in lactate dehydrogenase: a rescue pathway for maintaining redox balance, *J. Bacteriol. 181*, 5521–5526.

[214] Heux, S., Cachon, R. & Dequin, S. (2006) Cofactor engineering in *Saccharomyces cerevisiae*: Expression of a H_2O-forming NADH oxidase and impact on redox metabolism, *Metab. Eng. 8*, 303–314.

[215] Bennett, G.N., San, K.-Y. & Berrios-Rivera, S.J. (2002) The effect of increasing NADH availability on the redistribution of metabolic fluxes in *Escherichia coli* chemostat cultures, *Metab. Eng. 4*, 230–237.

[216] Sánchez, A.M., Bennett, G.N. & San, K.-Y. (2005) Effect of different levels of NADH availability on metabolic fluxes of *Escherichia coli* chemostat cultures in defined medium, *J. Biotechnol. 117*, 395–405.

[217] Ying, W. (2008) NAD^+/NADH and $NADP^+$/NADPH in cellular functions and cell death: Regulation and biological consequences, *Antioxidants & Redox Signaling 10*, 179–206.

[218] Stadtman, E.R., Cohen, G.N. & LeBras, G. (1961) Feedback inhibition and repression of aspartokinase activity in *Echerichia coli**, *Ann. N.Y. Acad. Sci. 94*, 952–959.

[219] Stifel, F.B. & Herman, R.H. (1972) Role of isozymes in metabolic control, *Am. J. Clin. Nutr. 25*, 606–611.

[220] Galili, G., Amir, R., Hoefgen, R. & Hesse, H. (2005) Improving the levels of essential amino acids and sulfur metabolites in plants., *Biol. Chem. 386*, 817–831.

[221] Galili, G. & Höfgen, R. (2002) Metabolic engineering of amino acids and storage proteins in plants, *Metab. Eng. 4*, 3–11.

[222] Moreno-Sánchez, R., Saavedra, E., Rodríguez-Enríquez, S. & Olín-Sandoval, V. (2008) Metabolic Control Analysis: A Tool for Designing Strategies to Manipulate Metabolic Pathways, *Journal of Biomedicine and Biotechnology 2008*, 597913.

[223] Maarleveld, T.R., Khandelwal, R.A., Olivier, B.G., Teusink, B. & Bruggeman, F.J. (2013) Basic concepts and principles of stoichiometric modeling of metabolic networks, *Biotechnol. J. 8*, 997–1008.

[224] Teusink, B., Bachmann, H. & Molenaar, D. (2011) Systems biology of lactic acid bacteria: a critical review., *Microb. Cell Fact. 10 Suppl 1*, S11–S11.

[225] Carbo, A., Hontecillas, R., Kronsteiner, B., Viladomiu, M., Pedragosa, M., Lu, P., Philipson, C.W., Hoops, S., Marathe, M., Eubank, S., Bisset, K., Wendelsdorf, K., Jarrah, A., Mei, Y. & Bassaganya-Riera, J. (2013) Systems modeling of molecular mechanisms controlling cytokine-driven CD4+ T cell differentiation and phenotype plasticity, *PLoS Comput. Biol. 9*, e1003027.

[226] Sarma, U. & Ghosh, I. (2012) Different designs of kinase-phosphatase interactions and phosphatase sequestration shapes the robustness and signal flow in the MAPK cascade, *BMC Syst. Biol. 6*, 82.

[227] Stanford, N.J., Lubitz, T., Smallbone, K., Klipp, E., Mendes, P. & Liebermeister, W. (2013) Systematic Construction of Kinetic Models from Genome-Scale Metabolic Networks, *PLoS One 8*, e79195.

[228] Mitchell, S. & Mendes, P. (2013) A Computational Model of Liver Iron Metabolism, *PLoS Comput. Biol. 9*, e1003299.

[229] Mc Auley, M.T., Wilkinson, D.J., Jones, J.J. & Kirkwood, T.B. (2012) A whole-body mathematical model of cholesterol metabolism and its age-associated dysregulation, *BMC Syst. Biol. 6*, 130.

[230] Boogerd, F., Bruggeman, F., Hofmeyr, J.-H. & Westerhoff, H. (2007) Towards philosophical foundations of systems biology: introduction, in *Systems Biology: Philosophical Foundations* (Boogerd, F., Bruggeman, F.J., Hofmeyr, J.-H.S. & Westerhoff, H.V., eds.), 1 edn., pp. 3–21, Elsevier, Amsterdam.

[231] Rohwer, J.M., Schuster, S. & Westerhoff, H.V. (1996) How to recognize monofunctional units in a metabolic system, *J. Theor. Biol. 179*, 213–228.

[232] Mashamaite, L.N., Rohwer, J.M. & Pillay, C.S. (2015) The glutaredoxin mono- and di-thiol mechanisms for deglutathionylation are functionally equivalent: implications for redox systems biology, *Biosci. Rep. 35*, e00173.

[233] Maxima (2016), Maxima 5.38.1 Manual: 14. Polynomials, http://maxima.sourceforge.net/docs/manual/maxima_14.html.

[234] Berlekamp, E.R. (1967) Factoring polynomials over finite fields, *The Bell System Technical Journal 46*, 1853–1859.

[235] SymPy Development Team (2016), GSoC-2016-Ideas: Multivariate polynomials and factorization, https://github.com/sympy/sympy/wiki/GSoC-2016-Ideas#multivariate-polynomials-and-factorization.

[236] Kronecker, L. (1882) Grundzüge einer arithmetischen Theorie der algebraischen Grössen, *J. Reine Angew. Math 92*, 1–122.

[237] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. & Kummer, U. (2006) COPASI—a COmplex PAthway SImulator, *Bioinformatics 22*, 3067–3074.

[238] Clewley, R. (2012) Hybrid Models and Biological Model Reduction with PyDSTool, *PLoS Computational Biology 8*, e1002628.